COGNITIVE SCIENCE

An Introduction to the Study of Mind

Jay Friedenberg ▪ Gordon Silverman

SAGE

SHOUT
SHARE
IT

COGNITIVE SCIENCE

COGNITIVE SCIENCE

An Introduction to the Study of Mind

Jay Friedenberg
Manhattan College

Gordon Silverman
Manhattan College

SAGE Publications
Thousand Oaks ▪ London ▪ New Delhi

Copyright © 2006 by Sage Publications, Inc.

All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

For information:

Sage Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

Sage Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

Sage Publications India Pvt. Ltd.
B-42, Panchsheel Enclave
Post Box 4109
New Delhi 110 017 India

Printed in the United States of America on acid-free paper.

Library of Congress Cataloging-in-Publication Data

Friedenberg, Jay.
Cognitive science : an introduction to the study of mind / Jay Friedenberg, Gordon Silverman.
p. cm.
Includes bibliographical references and indexes.
ISBN 1-4129-2568-1 (pbk.)
1. Cognitive science—Textbooks. I. Silverman, Gordon. II. Title.
BF311.F743 2006
153—dc22

2005009522

05 06 07 08 09 10 9 8 7 6 5 4 3 2 1

Acquiring Editor:	Jim Brace-Thompson
Associate Editor:	Margo Beth Crouppen
Production Editor:	Sanford Robinson
Editorial Assistant:	Karen Ehrmann
Typesetter:	C&M Digitals (P) Ltd.
Indexer:	Jeanne Busemeyer
Cover Designer:	Janet Foulger

Contents

Preface **xv**

Acknowledgments **xxiii**

1. Introduction: Exploring Inner Space **1**

A Brave New World 1
What Is Cognitive Science? 2
Representation 3
Digital Representations 6
Analog Representations 7
The Dual-Coding Hypothesis 7
Propositional Representations 8
Computation 9
The Tri-Level Hypothesis 10
The Classical and Connectionist Views of Computation 12
The Interdisciplinary Perspective 12
The Philosophical Approach 15
The Psychological Approach 16
The Cognitive Approach 17
The Neuroscience Approach 18
The Network Approach 18
The Evolutionary Approach 19
The Linguistic Approach 20
The Artificial Intelligence Approach 20
The Robotics Approach 21
In Depth: Categories of Mental Representation 22
Minds On Exercise: Analogical Reasoning 24

2. The Philosophical Approach: Enduring Questions **29**

What Is Philosophy? 29
The Mind-Body Problem 30
Flavors of Monism 33
Flavors of Dualism 34
Evaluating the Dualist Perspective 36
Functionalism 38
Evaluating the Functionalist Perspective 40
The Free Will–Determinism Debate 41
The Issue of Determinism 41
The Issue of Free Will 43
Evaluating the Free Will–Determinism Debate 44
The Knowledge Acquisition Problem 45
Evaluating the Knowledge Acquistion Debate 47
The Mystery of Consciousness 49
The What-It's-Like Argument 50
Mind as an Emergent Property 52
Evaluating the Emergent View of Mind 53
Consciousness and Neuroscience 54
Consciousness and Artificial Intelligence 56
Overall Evaluation of the Philosophical Approach 59
In Depth: Dennett's Multiple Drafts Theory of Consciousness 60
Minds On Exercise: Decision Making 63

3. The Psychological Approach: A Profusion of Theories **65**

What Is Psychology? 65
Psychology and the Scientific Method 67
Mental Atoms, Mental Molecules, and a Periodic Table of the Mind: The Voluntarist Movement 68
Evaluating the Voluntarist Approach 71
Structuralism: What the Mind Is 72
Evaluating the Structuralist Approach 73
Functionalism: What the Mind Does 74
Evaluating the Functionalist Approach 76
The Whole Is Greater Than the Sum of Its Parts: Mental Physics and the Gestalt Movement 77
Evaluating the Gestalt Approach 81
Mini-Minds: Mechanism and Psychoanalytic Psychology 82
Evaluating the Psychoanalytic Approach 84
Mind as a Black Box: The Behaviorist Approach 85
Evaluating the Behaviorist Approach 87

Overall Evaluation of the Psychological Approach 88
In Depth: Insight Learning 89
Minds On Exercise: Introspection 91

4. The Cognitive Approach I: History, Vision, and Attention 95

Some History First: The Rise of Cognitive Psychology 95
The Cognitive Approach: Mind as an Information Processor 97
Modularity of Mind 98
Evaluating the Modular Approach 99
Theories of Vision and Pattern Recognition 100
Template Matching Theory 100
Evaluating Template Matching Theory 101
Feature Detection Theory 103
Evaluating Feature Detection Theory 103
A Computational Theory of Vision 105
Evaluating the Computational Theory of Pattern Recognition 106
Feature Integration Theory 108
Evaluating Feature Integration Theory 111
Theories of Attention 112
Broadbent's Filter Model 113
Evaluating the Filter Model 115
Treisman's Attenuation Model 115
The Deutsch-Norman Memory Selection Model 116
The Multimode Model of Attention 116
Kahneman's Capacity Model of Attention 117
Evaluating the Capacity Model of Attention 119
Evaluating the Model-Building Approach 119
In Depth: Biederman's Recognition-by-Components Theory of Pattern Recognition 120
Minds On Exercise: Silhouettes and Object Constancy 122

5. The Cognitive Approach II: Memory, Imagery, and Problem Solving 125

Types of Memory 125
Sensory Memory 126
Working Memory 127
Long-Term Memory 131
Memory Models 133
The Modal Model 134
Evaluating the Modal Model 135

The ACT* Model 135
Evaluating the ACT* Model 137
The Working Memory Model 137
Evaluating the Working Memory Model 139
Visual Imagery 139
The Kosslyn and Schwartz Theory of Visual Imagery 140
Image Structures 141
Image Processes 142
Evaluating the Kosslyn and Schwartz Theory 146
The Imagery Debate 147
Problem Solving 149
The General Problem Solver Model 153
Evaluating the General Problem Solver Model 154
The SOAR Model 155
Evaluating the SOAR Model 156
Overall Evaluation of the Cognitive Approach 158
In Depth: Search in Working Memory 159
Minds On Exercise: Memory Effects 161

6. The Neuroscience Approach: Mind As Brain 163

The Neuroscience Perspective 163
Methodology in Neuroscience 164
Techniques for the Study of Brain Damage 164
Evaluating Techniques for the Study of Brain Damage 165
Brain Recording Techniques 166
Computerized Axial Tomography (CAT) 167
Positron Emission Tomography (PET) 167
Magnetic Resonance Imaging (MRI) 168
The Electrical Stimulation Technique 169
The Small Picture: Neuron Anatomy and Physiology 170
The Big Picture: Brain Anatomy 170
Directions in the Nervous System 172
The Cortex 173
The Split Brain 174
The Neuroscience of Visual Object Recognition 175
Visual Agnosias 176
Apperceptive Agnosia 177
Associative Agnosia 178
Face Perception 180

The Neuroscience of Attention 181
Models of Attention 185
A Component Process Model 185
Distributed Network Models 186
The Neuroscience of Memory 187
Learning and Memory 188
The Hippocampal System 198
Models of Hippocampal Function 190
Neural Substrates of Working Memory 192
Evaluating the Neuroscience of Working Memory 195
Neural Substrates of Long-Term Memories 195
The Neuroscience of Executive Function and Problem Solving 197
Theories of Executive Function 199
Overall Evaluation of the Neuroscience Approach 201
In Depth: Binding and Neural Synchrony 202
Minds On Exercise: Neural Functions 203

7. Outline: The Network Approach: Mind as a Web 207

The Network Perspective 207
Principles Underlying Artificial Neural Networks 208
Characteristics of Artificial Neural Networks 210
Early Conceptions of Neural Networks 212
Back Propagation and Convergent Dynamics 215
Artificial Neural Network Typologies 219
Evaluating the Connectionist Approach 221
Advantages 221
Problems and Disadvantages 222
Semantic Networks: Meaning in the Web 224
Characteristics of Semantic Networks 225
A Hierarchical Semantic Network 226
Evaluating the Hierarchical Model 228
Propositional Semantic Networks 230
Evaluating Semantic Networks 231
Overall Evaluation of the Network Approach 233
In Depth: NETtalk 234
Minds On Exercise: Free Association 236

8. The Evolutionary Approach: Change Over Time 239

The Evolutionary View 239
Evolutionary Psychology 240
A Little Background: Selection 241

Evolved Psychological Mechanisms 244
Evolution and Cognitive Processes 247
Categorization 247
Memory 248
Logical Reasoning 250
Judgment Under Uncertainty 253
Language 256
Sex Differences in Cognition 257
Evolutionary Computing 261
Artificial Life 263
Neural Darwinism 265
Evaluating Evolutionary Psychology 266
Overall Evaluation of the Evolutionary Approach 269
In Depth: A Stage Theory of Evolution 270
Minds On Exercise: Memory for Object Location 272

9. The Linguistic Approach: Language and Cognitive Science 275

The Linguistic Approach: The Importance of Language 275
The Nature of Language 276
Language Use in Primates 278
Evaluating Language Use in Primates 280
Language Acquisition 282
Evaluating Language Acquisition 284
Language Deprivation 284
Evaluating Language Deprivation 287
Philosophy and Linguistics: The Linguistic Relativity Hypothesis 288
Evaluating the Linguistic Relativity Hypothesis 290
Cognition and Linguistics: The Role of Grammar 291
Evaluating Universal Grammar 294
Neuroscience and Linguistics: The Wernicke-Geschwind Model 294
Evaluating the Wernicke-Geschwind Model 297
Artificial Intelligence and Linguistics: Natural Language Processing 298
Speech Recognition 299
Syntactic Analysis 301
Semantic Analysis 301
Pragmatic Analysis 302
Evaluation of Natural Language Processing 303
Overall Evaluation of the Linguistic Approach 304

In Depth: The Logogen Model of Word Recognition 304
Modern Conceptions of the Lexicon 307
Minds On Exercise: Conversational Pragmatics 308

10. Artificial Intelligence I: Definitional Perspective 311

Introduction 311
Historical and Philosophical Roots 312
The Quest for "Mechanical Life" 313
Philosophical Origins—Man as a Machine 315
Evaluating Descartes' Approach 316
Mechanical Computation 316
Defining Artificial Intelligence (AI) 320
Evaluating the Concept of AI 322
Strong AI 322
Applied AI 323
Cognitive Simulation and Natural Language Communication 323
AI Methodologies 331
The Computer as the Tool of AI Research 333
Evaluation of the Computer as a Model of Brain Organization 334
Programming 335
Evaluation of Programming Languages 336
Alan Turing and the Great Debate 336
Evaluation of the Turing Test (TT) and Turing's Detractors 341
Battle Lines: The Future of the TT 341
The TT as an Operational Definition of AI 341
The TT and Behaviorism 344
Is the TT Too Easy or Too Difficult? 344
Overall Evaluation of the AI Concept: Summarizing the Meaning of AI 345
In Depth: Behaviorism and Ned Block 347
Evaluating the Block Approach 349
Minds On Exercise: Play the IG Game 350

11. Artificial Intelligence II: Operational Perspective 353

Introduction 353
The Practical World of Artificial Intelligence 354
Goals of Practical Machines 354
Approaches to the Design of Intelligent Agents 356

Machine Intelligence, Knowledge,
and Machine Reasoning 357
Machine Representation of Knowledge 357
The Cyc Project 357
Evaluation of the Cyc Project 359
Objectives of Knowledge Representation 359
Characteristics of Knowledge Representation 360
Knowledge Representation Technologies 362
Predicates 362
Semantic Networks 364
Frames 365
Cases 366
Scripts 366
Machine Reasoning 367
Predicate Calculus 367
Logical Reasoning (Deduction, Abduction, Induction) 370
Drawing Inferences 372
Inductive Reasoning 379
Evaluation of Rule-Based Reasoning 379
Expert Systems 380
Expert Systems in Action 380
MYCIN 383
Evaluations of Expert Systems 384
Fuzzy Logic 386
Representation of Information in the Fuzzy World 387
Fuzzy Logic Rules 389
Fuzzy Advice for Decision-Making or Management 390
Evaluation of Fuzzy Logic Systems 390
Artificial Neural Nets (ANNs) 392
Overall Evaluation of the Operational Perspective 393
In Depth: The ID3 Algorithm
Minds On Exercise: Decision Making 396

12. Robotics: The Ultimate Intelligent Agents 399

Introduction 399
Some Robotic Achievements 402
Evaluating Robotic Potentials 404
Biological and Behavioral Foundations of Robotic Paradigms 404
Evaluation of the Lorenz/Tinbergen Approach 409
Transferring Behavioral Models to Robotics 410
Evaluation of the Biological Basis of Robotics 411

Foundations of Robotic Paradigms 412
Evaluation of Paradigm Foundations 413
Robotic Paradigms 414
Hierarchical Paradigm 414
Evaluation of the Hierarchical Paradigm 415
The Reactive Paradigm 416
Evaluation of the Reactive Paradigm 419
The Hybrid Deliberative/Reactive Paradigm 420
Evaluation of Hybrid Architectures 423
Overall Evaluation of Robots as Ultimate Intelligent Agents 425
In Depth: Autonomous Robot Architecture (AuRA) 426
Evaluation of AuRA 429
Minds On Exercise: Relational Graphs 429

13. Conclusion: Where We Go From Here 433

The Benefits of Cognitive Science 433
Working Memory: An Example of an Integrated Program of Study 434
The Philosophical Approach 435
The Psychological Approach 435
The Cognitive Approach 435
The Neuroscience Approach 437
The Network Approach 437
The Evolutionary Approach 438
The Linguistic Approach 438
The Artificial Intelligence Approach 438
The Robotics Approach 439
Issues in Cognitive Science 439
A General Issue: Lack of a Single Unified Theory 439
Specific Issues Facing Cognitive Science 440
Emotions 441
Consciousness 443
Physical Environments 444
Social Environments 446
Individual and Cultural Differences 448
Enhancing Cognitive Science 449
Evaluating Cognitive Science Theories 449
The Role of Integration 452
Integration Across Levels of Description 452
Integration Across Disciplines 453
Integration Across Methodologies 454

The Future 454
In Depth: Multiagent Systems 455
Minds On Exercise: Evaluating Theories of Mind 457

Glossary **459**

References **483**

Name Index **505**

Subject Index **513**

About the Authors **531**

Preface

No single volume could hope to describe, even succinctly, the tremendous number of discoveries in cognitive science. We do not attempt this. Our goal is to give readers a solid preparatory understanding of the major theoretical contributions of each cognitive science discipline. So, rather than exhaustively catalog all possible findings, we describe the major theories of mind that have been developed within each discipline. But what is a theory of mind? We define it as a unique, organized set of ideas that teaches us how to think about the mind. We also describe models of mind. Whereas a theory is general and may not be tied to a particular process, a model is specific and details how information is represented and computed by a particular process.

This perspective means that we focus on theories and models rather than the many experiments that have been conducted within cognitive science disciplines. We do not adopt the structure that is typical of textbooks in cognitive psychology, for example, one that is ordered around detailed methodological descriptions of classic experiments. We instead discuss general methods unique to a perspective at the outset of each chapter. Some experimental results are also presented, but these are intended to illustrate characteristics of a particular theory or model. Experiments and method in this book are thus used to service larger conceptual ideas of mind rather than the other way around.

The Contents of the Book

Each chapter (or set of chapters) is devoted entirely to a single disciplinary approach. These approaches include the evolutionary approach and robotics—topics not usually addressed in other texts. The general structure of each chapter is as follows: First, an approach is described in terms of what makes it

special. The major ideas that motivate each perspective and the problems each attempts to solve are laid out. Following this, we present factual background information that we believe is important and describe the approach's methodology. The bulk of each chapter is devoted to detailing the specific discipline's significant theories and models of mind. Where relevant, we evaluate the theory and outline its strengths and weaknesses. Each chapter ends with a section entitled "Overall Evaluation." These final sections evaluate the overall contributions of the approaches. We hope that these evaluations help students to realize that there is no single theory or approach that best captures what the mind is about, and to think critically about these ideas on their own.

Throughout there is a consistent set of topics that we touch upon. These topics are the major categories of mental processes. They include visual pattern or object recognition, attention, memory, imagery, and problem solving. Judgment and reasoning are dealt with primarily from an evolutionary perspective, while language has its own chapter. In some cases, these topics are introduced in the same order within chapters for ease of comparison. To further facilitate comparisons across chapters we have included a matrix (pp. xviii–xxii) that details the primary and secondary topics, methodologies, major figures, and an evaluation summary. The matrix is a useful tool for discerning similarities and differences across disciplines.

There are a number of distinctive features in this book. Every chapter concludes with supplementary material, exercises, and activities. In Depth sections illustrate specific theories in greater detail. The Minds On Exercises prompt students to work on group activities that will help them to further connect with chapter content. We also present Food for Thought sections with discussion questions drawn from material throughout the chapter. Some of these discussion questions are web questions that ask students to perform a specific activity or exercise on the Internet. Last but not least, there are lists of suggested readings. These readings are for the most part accessible to an introductory audience.

Beyond the Book

The text comes with many ancillary materials. Instructor's Resources that accompany the book are available on CD. The disk contains chapter outlines, PowerPoint lectures, and a test bank with different question types. We also include projects for instructors. These projects again encourage in-depth exploration of selected topics. Some of the projects are web-based, others require research, and all can serve as the basis of a topic for a paper. Students are able

to access a Companion Website containing electronic flashcards of glossary terms. The website also allows students to take practice quizzes to assess their level of understanding, and provides them with links used throughout the book and some additional sites for further exploration.

The Matrix

Chp.	*Name/Title*	*Chapter/Approach Summary*	*Primary Topic/Issues*	*Secondary Topic/Issues*	*Methodologies*	*Major Figures*	*Evaluation*
1	Introduction	An introduction to cognitive science and summary overview of different perspectives	What is cognitive science? Representation. Computation. The interdisciplinary perspective. Categories of mental representation.	Digital and analog representation. The dual-coding hypothesis. Concepts. Propositions. Production rules. Declarative and procedural knowledge. Analogies.	No methodologies discussed.	Thagard. Paivio. Pylyshyn. Marr.	Cognitive science is unique in that it binds together different perspectives and methodologies in the study of mind.
2	The Philosophical Approach	The search for wisdom and knowledge. Frames broad questions about mind.	The mind-body problem. Functionalism. Free will and determinism. Knowledge acquisition. Consciousness.	Monism. Dualism. Nature-nurture debate. Reductionism. Emergence.	Deductive and inductive reasoning.	Aristotle. Plato. Berkeley. Democritus. Descartes. Ryle. Clark. Hume. Rand. Locke. Chalmers. Nagel. Jackson. Searle. Churchland. Dennett.	Provides a broad perspective. Asks fundamental questions. Not an empirical approach.
3	The Psychological Approach	The scientific study of mind and behavior.	The scientific method. Voluntarism. Structuralism. Functionalism. Gestalt psychology. Psychoanalytic psychology. Behaviorism.	Theory and hypothesis. Independent and dependent variables. Experimental and control groups. Stream of consciousness. Levels of consciousness. Classical and operant conditioning.	Scientific method. Introspection. Phenomenology.	Wundt. Titchener. James. Wertheimer. Koffka. Kohler. Freud. Watson. Pavlov. Skinner.	Multiple theoretical positions. First systematic and scientific study of mental phenomena. Problems with introspection and phenomenology.

4	The Cognitive Approach I	The information-processing view of mind. Use of a computer as a metaphor for mind. Use of process models and assumption of modularity.	Information processing perspective. Modularity. Pattern recognition. Attention.	Template matching. Feature detection. Computational vision. Feature integration theory. Models of attention.	Experimentation. Modeling.	Neisser. Fodor. Selfridge. Norman. Marr. Treisman. Broadbent. Deutch. Posner. Snyder. Kahneman. Biederman.	Fruitful synergistic use of experimentation and model-building.
5	The Cognitive Approach II	The information-processing view of mind. Use of a computer as a metaphor for mind. Use of process models and assumption of modularity.	Memory. Models of memory. Visual imagery. Problem solving.	Memory types: sensory, working, and long-term. The modal, ACT*, and working memory models. The Kosslyn-Schwartz theory of visual imagery. Heuristics. Means-end analysis. The GPS and SOAR models. Memory scanning.	Experimentation. Modeling. Same as Cognitive Approach I chapter.	Sperling. Baddeley. Atkinson. Shiffrin. Anderson. Kosslyn. Block. Newell. Simon. Sternberg.	Common set of assumptions underlying information processing and modularity. Concepts of representation and computation need to be reconciled with connectionism.
6	The Neuroscience Approach	The study of nervous system anatomy and physiology that underlies and gives rise to cognitive function.	Neuroscience methodology. Neuron anatomy and physiology. Brain anatomy. Neuroscience of visual object recognition, attention, memory, executive function and problem solving.	The split brain. Dorsal and ventral pathways. Agnosias. Plasticity. Hippocampal function. Action schemas and scripts. Metacognition. Binding and neural synchrony.	Case studies. Lesion studies. Cell recording techniques. EEG, ERP, CAT, PET, and MRI. Electrical stimulation.	Sperry. Sacks. Humphreys. Posner. Mesulam. Lashley. Hebb. Shallice. Engel. Singer.	The marriage of cognitive and neuroscience perspectives in cognitive neuroscience is a good integrative approach. Specification of biological structures and processes of cognitive abilities.

(Continued)

Chp.	*Name/Title*	*Chapter/Approach Summary*	*Primary Topic/Issues*	*Secondary Topic/Issues*	*Methodologies*	*Major Figures*	*Evaluation*
7	The Network Approach	View of mind as an interconnected set of nodes or web. Processing consists of the spread of activation through the web.	Serial and parallel processing. Artificial neural networks. Semantic networks.	Perceptrons. Back propagation. Hopfield-Tank networks. Kohonen networks. Adaptive Resonance Theory. Stability and plasticity. Catastrophic interference. Spreading activation. Retrieval cues. Priming. Propositional networks.	Software simulations of artificial neural networks. Comparison of results with theory and empirical data.	McCulloch. Pitts. Hopfield. Kohonen. Grossberg. Collins. Quillian. Rumelhart. McClelland.	Significant advantages to using networks for understanding learning and knowledge representation. Challenges in building networks that rival the brain. Incompatibility of parallel processing, use of processing units and knowledge vs. behavior based approaches in comparison to traditional cognitive view.
8	The Evolutionary Approach	Mind as the adapted product of selection forces.	Natural selection. Evolved psychological mechanisms. Evolution and cognitive processes. Evolutionary computing. Artificial Life. Neural Darwinism.	General purpose vs. domain specific view of mind. Wason selection task. Heuristics and fallacies. Exaptation, molecular drive, and spandrels. Mosaic and regulatory models of development.	Experimentation. Comparison of species, individuals, and sexes.	Darwin. Buss. Cosmides. Tooby. Edelman. Donald.	Powerful theoretical framework. But not all mental processes may be adaptive. Good integration with neuroscience. Domain specific processing view clashes with general purpose processor view.

9	The Linguistic Approach	The multidisciplinary study of language.	The nature of language. Primate language use. Language acquisition. Language deprivation. Linguistic relativity hypothesis. Grammar. The Wernicke-Geschwind model. Natural language processing.	Phonology and morphology. Syntax and semantics. Animal language studies. Critical period. Second-language acquisition. Phrase structure, transformational, and universal grammar. Aphasias. Speech recognition. Pragmatic analysis.	Case studies. Network models. Developmental studies. Experimentation.	Gardner. Premack. Savage-Rumbaugh. Sapir. Whorf. Chomsky. Broca. Wernicke.	Multiple perspectives and techniques brought to bear on the complex cognitive topic of language. Advances in computer-based language comprehension.
10	Artificial Intelligence I	Defining the concept of Artificial Intelligence (AI). Machine representation of cognitive function.	Turing Test. Defining AI and its categories: strong AI; applied AI; cognitive simulation. Learning, reasoning, problem solving.	Automaton history. Philosophical origins of AI. Mechanical computation. Computer architecture as a cognitive model.	Machine models and simulation.	Turing. Babbage. Weizenbaum. Minsky. Papert. Block.	Historical streams of technology and philosophy merge to form the basis of AI. Definitions of AI depend on underlying interest; engineering to solve problems, cognitive science to codify knowledge and explain intelligence.

(Continued)

Chp.	*Name/Title*	*Chapter/Approach Summary*	*Primary Topic/Issues*	*Secondary Topic/Issues*	*Methodologies*	*Major Figures*	*Evaluation*
11	Artificial Intelligence Approach II	Operational perspective. Machine intelligence and reasoning.	Expert Systems. Fuzzy Logic. Neural Nets. Rule-based reasoning.	Logical Reasoning. Forward, backward reasoning. Searching methods.	Modelling. Simulation. Experimentation. Inductive, deductive, abductive reasoning.	Zadeh. Haack. Sowa. Lenat.	Imitation of human reasoning depends on top-down organization (Expert Systems, Fuzzy systems) or bottom up architectures (connectionist models). All must represent objects and processes within the world
12	Robotics	Understanding how to build an autonomous robot with cognitive features.	Robotic Paradigms: Hierarchical; Reactive Deliberative/Reactive;	Biological basis of Robotics. Foundations of Robotic Paradigms: SENSE, PLAN, ACT	Case study. Modelling. Robotic architectures and organization.	Lorenz/Tinbergen. Brooks. Arkin.	PLAN-SENSE-ACT architectures will seem primitive in the near future. Using signals from brain, autonomous robots will achieve the objectives of intelligent agents.
13	Conclusion	An evaluation of the cognitive science approach.	Benefits of cognitive science. Issues facing cognitive science. Enhancing cognitive science. Multiagent systems.	Emotions. Consciousness. Physical and social environments. Individual and cultural differences.	No methodologies discussed.	Wilson. Thagard. Damasio. Dreyfus. Gibson. Gardner.	Benefits of cognitive science are many and widespread throughout engineering, medicine, education, and other fields. Lack of a single unified theory. Inadequate accounts of emotions, consciousness, physical and social environments, and individual and cultural differences.

Acknowledgments

We would like to acknowledge the efforts of Kelly Carpiniello and Kristi Ferraro, who contributed significantly to the creation of the test questions. Tom McDonagh was also a very helpful research assistant on this project. Our thanks extend additionally to Jim Brace-Thompson and Margo Beth Crouppen for their detailed and supportive feedback during the editing process.

1

Introduction: Exploring Inner Space

"The sciences have developed in an order the reverse of what might have been expected. What was most remote from ourselves was first brought under the domain of law, and then, gradually, what was nearer: first the heavens, next the earth, then animal and vegetable life, then the human body, and last of all (as yet very imperfectly) the human mind."

—Bertrand Russell, 1935

A Brave New World

We are in the midst of a revolution. For centuries science has made great strides in our understanding of the external observable world. Physics revealed the motion of the planets, chemistry discovered the fundamental elements of matter, biology has told us how to understand and treat disease. But during much of this time, there were still many unanswered questions about something perhaps even more important to us. That something is the human mind.

What makes mind so difficult to study is that, unlike the phenomena described above, it is not something we can easily observe, measure, or manipulate. In addition, the mind is the most complex entity in the known universe.

To give you a sense of this complexity consider the following. The human brain is estimated to contain ten billion to one hundred billion individual nerve cells or neurons. Each of these neurons can have as many as ten thousand connections to other neurons. This vast web is the basis of mind, and gives rise to all of the equally amazing and difficult-to-understand mental phenomena such as perception, memory, and language.

The past several decades have seen the introduction of new technologies and methodologies for studying this intriguing organ. We have learned more about the mind in the past half-century than in all the time that came before that. This period of rapid discovery has coincided with an increase in the number of different disciplines—many of them entirely new—that study mind. Since then, a coordinated effort among the practitioners of these disciplines has come to pass. This interdisciplinary approach has since become known as cognitive science. Unlike the science that came before, which was focused on the world of external, observable phenomena, or "outer space," this new endeavor turns its full attention now to the discovery of our fascinating mental world, or "inner space."

What Is Cognitive Science?

Cognitive science can be roughly summed up as the scientific interdisciplinary study of the mind. Its primary methodology is the scientific method, although as we will see, many other methodologies also contribute. A hallmark of cognitive science is its interdisciplinary approach. It results from the efforts of researchers working in a wide array of fields. These include philosophy, psychology, linguistics, artificial intelligence, robotics, and neuroscience. Each field brings with it a unique set of tools and perspectives. One major goal of this book is to show that when it comes to studying something as complex as the mind, no single perspective is adequate. Instead, intercommunication and cooperation among the practitioners of these disciplines tell us much more.

The term *cognitive science* refers not so much to the sum of all these disciplines but to their intersection or converging work on specific problems. In this sense, cognitive science is not a unified field of study like each of the disciplines themselves, but a collaborative effort among researchers working in the various fields. The glue that holds cognitive science together is the topic of mind and, for the most part, the use of scientific methods. In the concluding chapter, we talk more about the issue of how unified cognitive science really is.

In order to really understand what cognitive science is all about we need to know what its theoretical perspective on the mind is. This perspective centers

on the idea of **computation**, which may alternatively be called information processing. Cognitive scientists view the mind as an information processor. Information processors must both represent and transform information. That is, a mind, according to this perspective, must incorporate some form of mental representation and processes that act on and manipulate that information. We will discuss these two ideas in greater detail later in this chapter.

Cognitive science is often credited with being influenced by the rise of the computer. Computers are of course information processors. Think for a minute about a personal computer. It performs a variety of information-processing tasks. Information gets into the computer via input devices, such as a keyboard or modem. That information can then be stored on the computer, for example, on a hard drive or other disk. The information can then be processed using software such as a text editor. The results of this processing may next serve as output, either to a monitor or printer. In like fashion, we may think of people performing similar tasks. Information is "input" into our minds through perception—what we see or hear. It is stored in our memories and processed in the form of thought. Our thoughts can then serve as the basis of "outputs," such as language or physical behavior.

Of course this analogy between the human mind and computers is at a very high level of abstraction. The actual physical way in which data is stored on a computer bears little resemblance to human memory formation. But both systems are characterized by computation. In fact, it is not going too far to say that cognitive scientists view the mind as a machine or mechanism whose workings they are trying to understand.

Representation

As mentioned before, representation is fundamental to cognitive science. But what is a representation? Before listing the characteristics of a representation, it is helpful to describe briefly four categories of representation. A concept stands for a single entity or group of entities. Single words are good examples of concepts. The word "apple" denotes the concept of that particular type of fruit. Propositions are statements about the world and can be illustrated with sentences. The sentence "Mary has black hair" is a proposition that is itself made up of concepts. Rules are yet another form of representation that can specify the relationships between propositions. For example, the rule "If it is raining, I will bring my umbrella," makes the second proposition contingent on the first. There are also analog representations. An analogy helps us to make comparisons between two similar situations. We will discuss all four of

these representations in greater detail in the In Depth section at the end of this chapter.

There are four crucial aspects of any representation (Hartshorne, Weiss & Burks, 1931–1958). First, a "representation bearer" such as a human or a computer must realize a representation. Second, a representation must have content—meaning it stands for one or more objects. The thing or things in the external world that a representation stands for are called **referents**. A representation must also be "grounded." That is, there must be some way in which the representation and its referent come to be related. Fourth, a representation must be interpretable by some interpreter, either the representation bearer him or herself, or somebody else. These and other characteristics of representations are discussed next.

The fact that a representation stands for something else means it is **symbolic**. We are all familiar with symbols. We know for instance that the symbol "$" is used to stand for money. The symbol itself is not the actual money, but instead is a surrogate that refers to its referent, which is actual money. In the case of mental representation, we say there is some symbolic entity "in the head" that stands for real money. Figure 1.1 shows some aspects of a mental representation of money. Mental representations can stand for many different types of things and are by no means limited to simple conceptual ideas such as "money." Research suggests that there are more complex mental representations that can stand for rules, for example, knowing how to drive a car, and analogies, which may enable us to solve certain problems or notice similarities (Thagard, 2000). See the In Depth section for a more detailed discussion of these other forms of mental representation.

Human mental representations, especially linguistic ones, are said to be **semantic**, which is to say they have meaning. Exactly what constitutes meaning and how a representation can come to be meaningful are topics of debate. According to one view, a representation's meaning is derived from the relationship between the representation and what it is about. The term that describes this relation is **intentionality**. Intentionality means "directed upon an object." Mental states and events are intentional. They refer to some actual thing or things in the world. If you think about your brother, then the thought of your brother is directed toward him, not toward your sister, a cloud, or some other object.

Intentionality is considered to have at least two properties. The first is **isomorphism**, or similarity of structure between a representation and its referent. This similarity means one can map different aspects of a representation onto its referent. Analog visual images, discussed further below, are good examples of this property. This is because they are believed to preserve the spatial

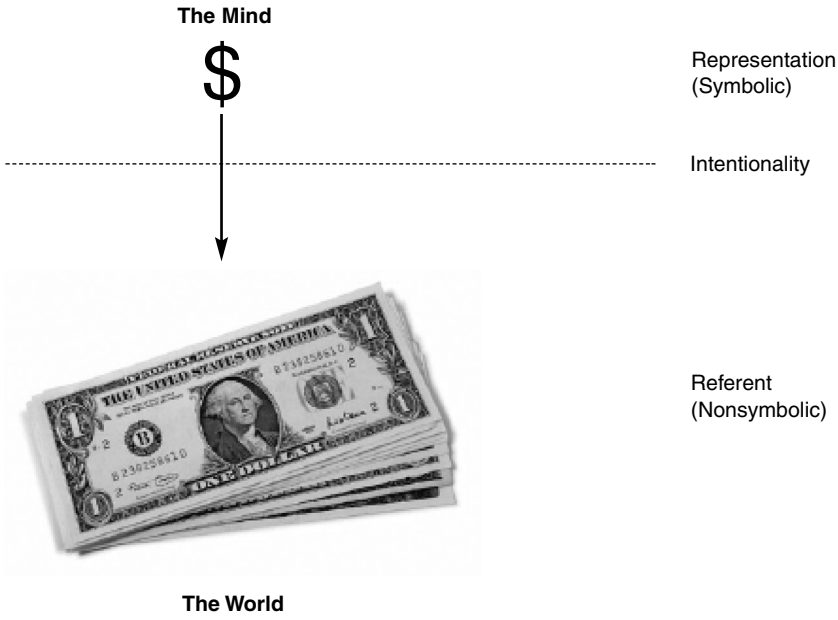
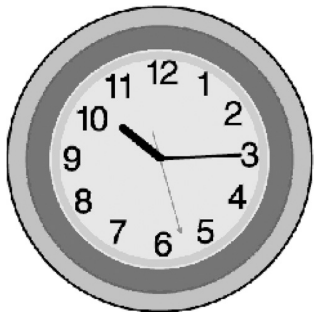


Figure 1.1 Different aspects of the symbolic representation of money

characteristics of the referent. A visual image of a cruise ship, for instance, would have greater horizontal than vertical extent because these boats are much longer than they are tall. The researcher Stephen Kosslyn has shown that it takes longer to "scan" a visual image across a dimension where distances between points in the object are greater and relatively less time across a dimension where such distances are shorter. The section on visual imagery contains more on the methods and results of this experiment and others that demonstrate the isomorphic characteristics of images.

A second characteristic of intentionality has to do with the relationship between inputs and outputs to the world. An intentional representation must be triggered by its referent or things related to it. Consequently, activation of a representation (i.e., thinking about it) should cause behaviors or actions that are somehow related to the referent. For example, if your friend Sally told you about a cruise she took around the Caribbean last December, an image of a cruise ship would probably pop to mind. This might then cause you to ask her if the food on board was good. Sally's mention of the cruise was the stimulus input that activated the internal representation of the ship in your mind. Once

A digital representation of time



An analog representation of time

Figure 1.2 Digital and analog clocks represent time in fundamentally different ways

it was activated, it caused the behavior of asking about the food. This relation between inputs and outputs is known as an **appropriate causal relation**.

Digital Representations

In a **digital representation**, sometimes also known as a symbolic representation, information is coded in a discrete way with set values. A digital clock for example, represents time discretely (see Figure 1.2). It displays a separate number for each hour, minute, or year. There are distinct advantages to digital representations. They specify values exactly. The symbols used in digital representations, such as numbers, can be operated on by a more general set of processes than analog structures. In mathematics, a wide range of operators such as addition, division, or squaring can be applied to digital number representations. The results of these operations are new numbers that can themselves be transformed by additional operations.

Language can serve as an example of a digital mental representation, and in fact verbal concepts seem to be the system of human symbolic representation that is most commonly used. The basic elements of written language are letters. These are discrete symbols that are combined according to a set of rules. The combinations, or words, have meaning and are themselves combined into other higher-order units, sentences, which also have semantic content. The rules by which these word elements are combined and transformed in language are called **syntax**. Syntax constitutes the set of permissible operations on the word elements. It is the elements themselves that are the mental representations. In

the chapter on linguistics, we talk more about linguistic representations and syntax.

Analog Representations

Analog representations, in contrast, represent information in a continuous way. Information in an analog system can theoretically take on any value not limited by resolution. Resolution refers to the amount of detail contained in an analog representation. Representations with high resolution have correspondingly more information. An analog clock represents time through the movement of its various hands. The positions of these hands on the dial indicate the time (see Figure 1.2). In addition to being able to represent large numbers of values, analog representations have the advantage of providing simple, direct solutions to some problems. They do however have a greater computational margin of error and because of the smaller number of operations that can be performed on them, are more limited for use in problem-solving.

Visual images are the best example of mental analog representations. Researchers in cognitive psychology have conducted numerous experiments that strongly suggest we represent visual information in an analog fashion. Stop reading for a moment and close your eyes. Imagine a picture of a palm tree on a sunny beach. Can you see the pattern on the bark? What about the coconuts? Images capture many of the same properties as their referents, such as distances between corresponding sets of points. The types of transformations that can be performed on images are also the kinds of changes that physical objects in the external observable world undergo. These include rotations, translations, and reflections. In the section on visual imagery in the cognitive chapter, we elaborate on the nature of visual images and discuss experiments that reveal the kinds of operations that can be performed on them.

The Dual-Coding Hypothesis

The use of both digital/symbolic and image representations collectively has been referred to as the **dual-code hypothesis** (Paivio, 1971). Alan Paivio believes that many ideas can be represented in either of these two forms interchangeably. This is especially true for a specific concrete concept, such as "elephant," for which we can form a visual image or a verbal representation. However, there are some concepts for which a symbolic code seems more appropriate. Take the idea of "justice." This is abstract, and although we could attach an image to it, such as that of a court building, there is no unambiguous and unique identifying image.

Evidence in support of dual-code theory comes from studies in which better recall is demonstrated for words representing concrete concepts, as compared to words representing abstract concepts (Paivio, 1971). According to Paivio, the reason for this is that two codes are better than one. Let's assume a subject in a memory experiment is presented with the word "elephant" and forms two codes to remember it. If he or she has forgotten one code later on at recall, he or she should still be able to access and retrieve the other. In this case, the image of the elephant may come to mind even if its symbolic word representation has faded.

Propositional Representations

Propositions are a third major category of representation, in addition to symbolic and imaginal codes (Pylyshyn, 1973). According to the **propositional hypothesis**, mental representations take the form of abstract sentence-like structures. Propositions are good at capturing the relationships between concepts. For example, the sentence "Mary looked at John" specifies a type of relationship between Mary and John, and that relationship can then be translated into either a verbal symbolic code, as in the actual form of a sentence, or an image code.

Propositions are believed to lie in a deep format that is neither visual nor verbal. This format can best be described as a logical relationship among constituent elements and is denoted by a **predicate calculus**. A predicate calculus is a general system of logic that accurately expresses a large variety of assertions and modes of reasoning. The proposition "Mary looked at John" can be represented by a predicate calculus such as:

[Relationship between elements] ([Subject element], [Object element])

where "Mary" is the subject element, "John" is the object element, and "looking" is the relationship between elements. What is nice about a predicate calculus is that it captures the essential logical structure of a complex idea independent of its actual elements. Any number of subjects, objects, and relationships can be inserted into the abstract format of a proposition. A proposition is thus believed to capture the basic meaning of a complex idea. This basic meaning, when translated back into a symbolic or visual code, can then be expressed in a variety of ways. For example, the sentences "Mary looked at John" and "John was looked at by Mary" are two alternate verbal codes for the same proposition. Likewise, one could form several different visual images to convey the one proposition.

Although a predicate calculus is a nice way of expressing a proposition, it doesn't mean that the proposition actually assumes this format in our brains. In fact, it is not clear exactly how propositions are mentally instantiated or realized. They do, however, serve as very useful hypothetical constructs because they are concise and can specify virtually all of the possible relationships between concepts.

To sum up this section, mental representations are powerful. They allow for the creation of an inner world that we can think about. The byproducts of these thoughts allow us to understand and interact successfully with the environment. Rather than knocking about in the world and making mistakes or taking risks, we can use representations to plan and carry out appropriate actions. Furthermore, the formal implementation of representations in a set of symbols, such as we envisage in mental pictures or language, allows us to communicate our thoughts to others. This in turn gives rise to more complex and adaptive forms of social cooperation.

Computation

As mentioned earlier, representations are only the first key component of the cognitive science view of mental processes. Representations by themselves are of little use unless something can be done with them. Having the concept of money doesn't do much for us unless we know how to calculate a tip or can give back the correct amount of change to someone. In the cognitive science view, the mind performs computations on representations. It is therefore important to understand how and why these mental mechanisms operate.

What sorts of mental operations does the mind perform? If we wanted to get detailed about it, the list would be endless. Take the example of mathematical ability. If there were a separate mental operation for each step in a mathematical process, we could say the mind adds, subtracts, divides, and so on. Likewise, with language we could say there are separate mental operations for making a noun plural, putting a verb into past tense, and so on. It is better, then, to think of mental operations as falling into broad categories. These categories can be defined by the type of operation that is performed or by the type of information acted upon. An incomplete list of these operations would include sensation, perception, attention, memory, language, mathematical reasoning, logical reasoning, decision making, and problem-solving. Many of these categories may incorporate virtually identical or similar sub-processes, for example, scanning, matching, sorting, and retrieving. Figure 1.3 shows the kinds of mental processes that may be involved in solving a simple addition problem.

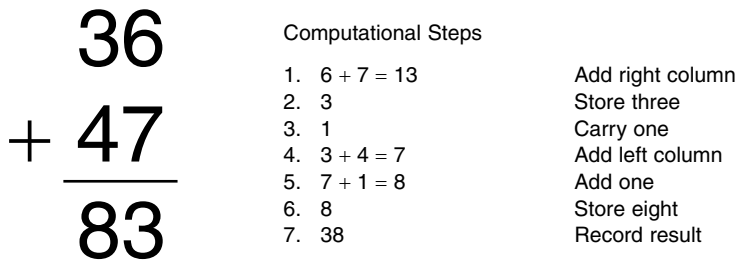


Figure 1.3 Some of the computational steps involved in solving an addition problem

The Tri-Level Hypothesis

Any given information process can be described at several different levels. According to the tri-level hypothesis, mental or artificial information-processing events can be evaluated on at least three different levels (Marr, 1982). The highest or most abstract level of analysis is the **computational level**. At this level, one is concerned with two tasks. The first is a clear specification of what the problem is. Taking the problem as it may have originally been posed, in a vague manner perhaps, and breaking it down into its main constituents or parts can bring about this clarity. It means describing the problem in a precise way such that the problem can be investigated using formal methods. It is like asking the questions: What exactly is this problem? What does this problem entail? The second task one encounters at the computational level concerns the purpose or reason for the process. The second task consists of asking: Why is this process here in the first place? Inherent in this analysis is the idea of adaptiveness—the idea that human mental processes are learned or have evolved to enable the human organism to solve a problem it faces. This is the primary explanatory perspective used in the evolutionary approach. We describe a number of cognitive processes and the putative reasons for their evolution in the chapter devoted to that approach.

Stepping down one level of abstraction, we can next inquire about the actual way in which an information process is carried out. To do this we need an **algorithm**, a formal procedure or system that acts on informational representations. It is important to note that algorithms can be carried out regardless of a representation's meaning; algorithms act on the form, not the meaning, of the symbols they transform. One way to think of algorithms is that they are "actions" used to manipulate and change representations. Algorithms are formal, meaning they are well-defined. We know exactly what occurs at each step of an

algorithm and how a particular step changes the information being acted on. A mathematical formula is a good example of an algorithm. A formula specifies how the data is to be transformed, what the steps are, and what the order of steps is. This type of description is put together at the **algorithmic level**, sometimes also called the programming level. It is equivalent to asking the question: What information-processing steps are being used to solve the problem? If we were to draw an analogy with computers, the algorithmic level is like software, because software contains instructions for the processing of data.

The most specific and concrete type of description is formulated at the **implementational level**. Here we ask: What is the information processor made of? What types of physical or material changes underlie changes in the processing of the information? This level is sometimes referred to as the hardware level, since, in computer parlance, the hardware is the physical "stuff" the computer is made of. This would include its various parts—a monitor, harddrive, keyboard, and mouse. At a smaller scale, computer hardware consists of circuits and even the flow of electrons through the circuits. The hardware in human or animal cognition is the brain and, at a smaller scale, the neurons and activities of those neurons.

At this point, one might wonder: Why do we even need an algorithmic or formal level of analysis? Why not just map the physical processes at the implementational level onto a computational description of the problem, or alternatively, onto the behaviors or actions of the organism or device? This seems simpler, and we need not resort to the idea of information and representation. The reason is that the algorithmic level tells us how a particular system performs a computation. Not all computational systems solve a problem in the same way. Computers and humans can both perform addition, but do so in drastically different fashions. This is true at the implementational level obviously, but understanding the difference formally tells us much about alternative problem-solving approaches. It also gives us insights into how these systems might compute solutions to other novel problems that we might not understand.

This partitioning of the analysis of information-processing events into three levels has been criticized as being fundamentally simplistic, since each level can in turn be further subdivided into levels (Churchland, Koch & Sejnowski, 1990). Figure 1.4 depicts one possible organization of the many structural levels of analysis in the nervous system. Starting at the top, we might consider the brain as one organizational unit; brain regions as corresponding to another organizational unit one step down in spatial scale; and then neural networks, individual neurons, and so on. Similarly, we could divide algorithmic steps into different sub-steps, and problems into sub-problems. To compound all this, it is not entirely clear how to map one level of analysis onto another.

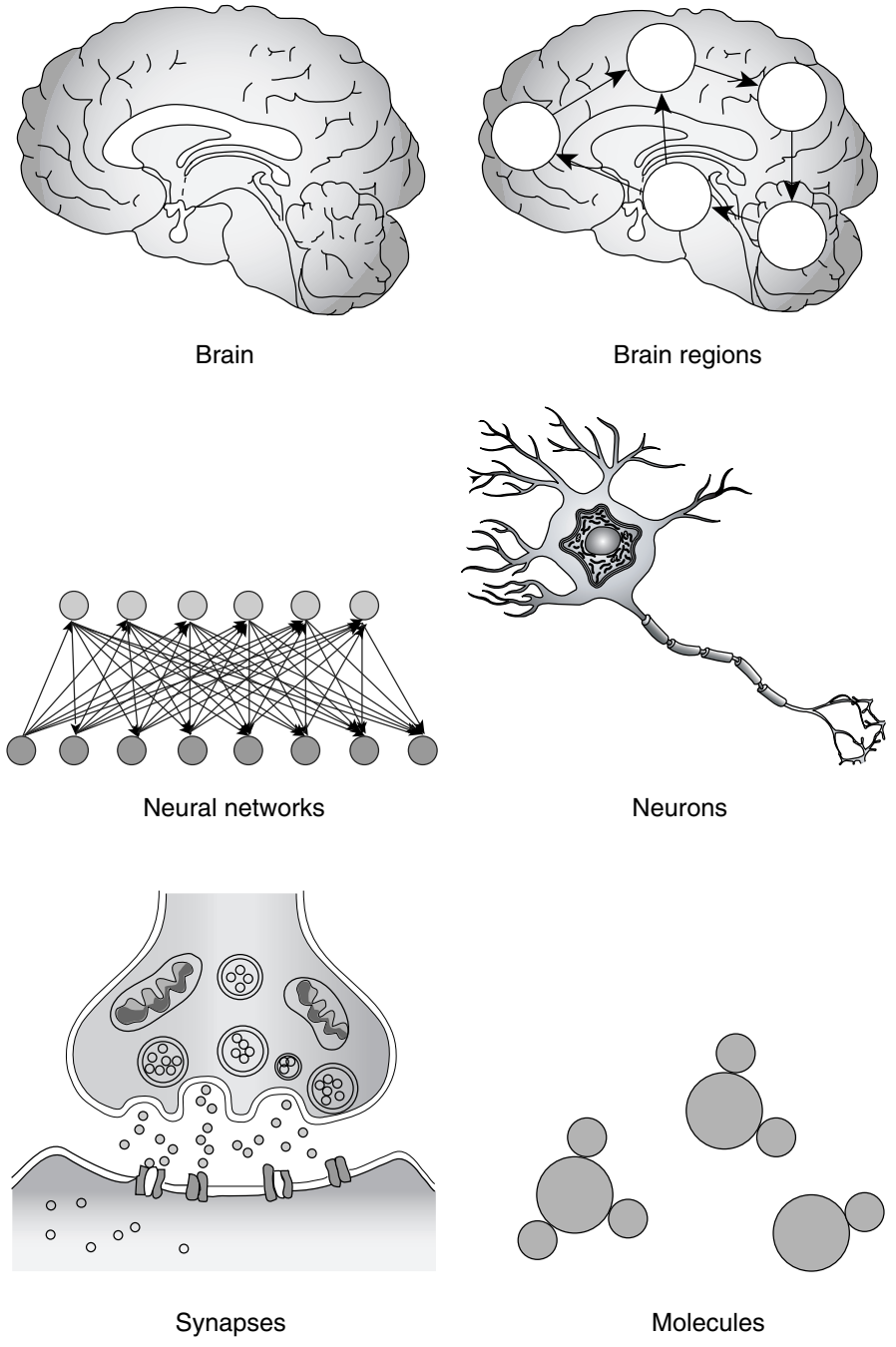


Figure 1.4 Structural levels of analysis in the nervous system

We may be able to clearly specify how an algorithm executes, but be at a loss to say exactly where or how this is achieved with respect to the nervous system.

The Classical and Connectionist Views of Computation

Before finishing our discussion of computation, it is important to differentiate between two distinct conceptions of what it is. So far, we have been talking about computation as being based on the formal systems notion. In this view a computer is a **formal symbol manipulator**. Let's break this definition down into its component parts. A system is formal if it is syntactic or rule-governed. The rules of language and mathematics are formal systems because they specify which types of allowable changes can be made to symbols. Formal systems also operate on representations independent of the content of those representations. In other words, a process can be applied to a symbol regardless of its meaning or semantic content. A symbol, as we have already indicated, is a form of representation and can assume a wide variety of forms. Manipulation here implies that computation is an active, embodied process that takes place over time. That is, manipulations are actions, they occur physically in some type of computing device, and they take some time to occur, that is, they don't happen instantaneously.

But this is not the only conception of what computation is. The network approach to computation differs from the classical formal systems approach in cognitive science in several ways. In the classical view, knowledge is represented locally, in the form of symbols. In the connectionist view knowledge is represented as a pattern of activation or weights that is distributed throughout a network. Processing style is also different in each approach. The classical view has processing occurring in discrete stages, whereas in connectionism, processing occurs in parallel through the simultaneous activation of nodes. Some cognitive scientists downplay these differences, arguing that information processing occurs in both systems and that the tri-level hypothesis can be applied equally to both (Dawson, 1998). We compare and contrast the classical and connectionist views at the beginning of the network approach chapter.

The Interdisciplinary Perspective

There is an old fable about five blind men who stumble upon an elephant (see Figure 1.5). Not knowing what it is, they start to feel the animal. One man feels only the elephant's tusk and thinks he is feeling a giant carrot. A second man,

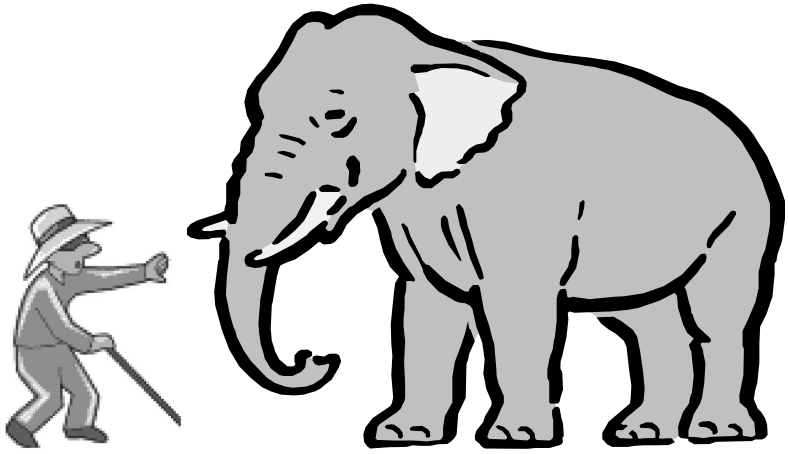


Figure 1.5 If you were the blind man, would you know it is an elephant?

feeling the ears, believes the object is a big fan. The third feels the trunk and proclaims it is a pestle, while a fourth touching only the leg believes it is a mortar. The fifth man, touching the tail, has yet another opinion: he believes it to be a rope. Obviously, all five men are wrong in their conclusions because each has only examined one aspect of the elephant. If the five men had gotten together and shared their findings, they may have easily pieced together what kind of creature it was. This story serves as a nice metaphor for cognitive science. We can think of the elephant as the mind and the blind men as researchers in different disciplines in cognitive science. Each individual discipline may make great strides in understanding its particular subject matter, but, if it cannot compare its results to those of other related disciplines, may miss out on understanding the real nature of what it is that is being investigated.

The key, then, to figuring out something as mysterious and complex as mind is communication and cooperation among disciplines. This is what's meant when one talks about cognitive science—not the sum of each of the disciplines or approaches, but their union. Recent years have seen an increase in this cooperation. A number of major universities have established interdisciplinary cognitive science centers, where researchers in such diverse areas as philosophy, neuroscience, and cognitive psychology are encouraged to work together on common problems. Each area can then contribute its unique strength to the phenomenon under study. The philosophers can pose broad questions and hypotheses, the neuroscientists can measure physiological performance and brain activity, while the cognitive psychologists can design and carry out experiments. The

consequent exchange of results and ideas then leads to fruitful synergies between these disciplines, accelerating progress with respect to finding solutions to the problem and yielding insights into other research questions.

We have alluded to some of the different approaches in cognitive science. Because this book is about explaining each approach and its major theoretical contributions, it is worth describing each now in terms of its perspective, history, and methodology. In the following sections we will also provide a brief preview of the issues addressed by each approach.

The Philosophical Approach

Philosophy is the oldest of all the disciplines in cognitive science. It traces its roots back to the ancient Greeks. Philosophers have been active throughout much of recorded human history, attempting to formulate and to answer basic questions about the universe. This approach is free to study virtually any sort of important question on virtually any subject, ranging from the nature of existence to the acquisition of knowledge, to politics, ethics, and beauty. Philosophers of mind narrow their focus to specific problems concerning the nature and the characteristics of mind. They might ask questions like: What is mind? How do we come to know things? How is mental knowledge organized?

The primary method of philosophical inquiry is reasoning, both deductive and inductive. **Deductive reasoning** involves the application of the rules of logic to statements about the world. Given an initial set of statements assumed to be true, philosophers can derive other statements that logically must be correct. For example, if the statement "College students study three hours every night" is true and the statement "Mary is a college student" is true, we can then conclude that "Mary will study three hours every night." Philosophers also engage in **inductive reasoning**. They make observations about specific instances in the world, notice commonalities among them, and draw conclusions. An example of inductive reasoning would be: "Whiskers the cat has four legs," "Scruffy the cat has four legs," therefore "All cats have four legs." However, philosophers do not use a systematic form of induction known as the scientific method. That is employed within the other cognitive science disciplines.

In Chapter 2, we summarize several of the fundamental issues facing philosophers of mind. With respect to the mind-body problem, philosophers wrangle over what exactly a mind is. Is the mind something physical like a rock or a chair, or is it nonphysical? Can minds exist only in brains or can they emerge from the operation of other complex entities such as computers? In

the free will–determinism debate we explore whether our actions can ever be completely known and/or predicted beforehand. The knowledge acquisition problem deals with how we come to know things. Is knowledge a product of one's genetic endowment or does it arise through one's interaction with the environment? How much does each of these factors contribute to any given mental ability? We also look into one of the most fascinating and enigmatic mysteries of mind, that of consciousness. What is consciousness? Are we really conscious at all?

The Psychological Approach

Compared to philosophy, psychology is a relatively young discipline. It can be considered to be old though, particularly when it is compared to some of the more recent newcomers to the cognitive science scene, for example, artificial intelligence and robotics. Psychology arose in the late 19th century and was the first discipline in which the scientific method was applied exclusively to the study of mental phenomena. Early psychologists established experimental laboratories that would enable them to catalog mental ideas and to investigate various mental capacities, such as vision and memory. Psychologists apply the scientific method to both mind and behavior. That is, they attempt to understand not just internal mental phenomena, such as thoughts, but also the external behaviors that these internal phenomena can give rise to.

The scientific method is a way of getting hold of valid knowledge about the world. One starts with a hypothesis or idea about how the world works and then designs an experiment to see if the hypothesis has validity. In an experiment, one essentially makes observations under a set of controlled conditions. The resulting data then either support or fail to support the hypothesis. This procedure, employed within psychology and cognitive science in general, is described more fully at the start of Chapter 3.

The field of psychology is broad and encompasses many subdisciplines, each one having its unique theoretical orientations. Each discipline has a different take on what mind is. The earliest psychologists, that is, the voluntarists and structuralists, viewed the mind as a kind of test tube in which chemical reactions between mental elements took place. In contrast, functionalism viewed mind not according to its constituent parts, but according to what its operations were—what it could do. The Gestaltists again went back to a vision of mind as composed of parts, but emphasized that it was the combination and interaction of the parts, which give rise to new wholes, that was important. Psychoanalytic psychology conceives of mind as a collection of

competing entities, while behaviorism sees it as a device that maps stimuli onto behaviors.

The Cognitive Approach

Starting in the 1960s a new form of psychology arrived on the scene. Known as cognitive psychology, it came into being in part as a backlash against the behaviorist movement and its profound emphasis on behavior. Cognitive psychologists placed renewed emphasis on the study of internal mental operations. They adopted the computer as a metaphor for mind, and described mental functioning in terms of representation and computation. They believed that the mind, like a computer, could be understood in terms of information processing.

The cognitive approach was also better able to explain phenomena such as language acquisition, for which behaviorists did not have good accounts. At around the same time, new technologies that allowed better measurement of mental activity were being developed. This promoted a movement away from the behaviorist's emphasis on external observable behaviors toward the cognitive scientist's emphasis on internal functions, as these could, for the first time, be observed with reasonable precision.

Inherent in the cognitive approach is the idea of modularity. Modules are functionally independent mental units that receive inputs from other modules, perform a specific processing task, and pass the results of their computation onto yet additional modules. The influence of the modular approach can be seen in the use of process models or flow diagrams. These depict a given mental activity via the use of boxes and arrows, where boxes depict modules and arrows the flow of information among them. The techniques used in this approach are the experimental method and computational modeling. Computational modeling involves carrying out a formal (typically software-based) implementation of a proposed cognitive process. Researchers can run the modeling process so as to simulate how the process might operate in a human mind. They can then alter various parameters of the model or change its structure in an effort to achieve results as close as possible to those obtained in human experiments. This use of modeling and comparison with experimental data is a unique characteristic of cognitive psychology and is also used in the artificial intelligence and network approaches.

Cognitive psychologists have studied a wide variety of mental processes. These include pattern recognition, attention, memory, imagery, and problem-solving. Theoretical accounts and processing models for each of these are given in Chapters 4 and 5. Language is within the purview of cognitive psychology,

but because the approach to language is also multidisciplinary, we describe it separately in Chapter 9.

The Neuroscience Approach

Brain anatomy and physiology have been studied for quite some time. Recent times however have seen tremendous advances in our understanding of the brain, especially in terms of how neuronal processes can account for cognitive phenomena. The general study of the brain and endocrine system is called neuroscience. The attempt to explain cognitive processes in terms of underlying brain mechanisms is known as cognitive neuroscience.

Neuroscience, first and foremost, provides a description of mental events at the implementational level. It attempts to describe the biological "hardware" upon which mental "software" supposedly runs. However, as discussed above, there are many levels of scale when it comes to describing the brain, and it is not always clear which level provides the best explanation for any given cognitive process. Neuroscientists, however, investigate at each of these levels. They study the cell biology of individual neurons and of neuron-to-neuron synaptic transmission, the patterns of activity in local cell populations, and the interrelations of larger brain areas.

A reason for many of the recent developments in neuroscience is, again, the development of new technologies. Neuroscientists employ a wide variety of machines to measure the performance of the brain at work. These include positron emission tomography (PET) scanners, computerized axial tomography (CAT) scanners, and magnetic resonance imaging (MRI) machines. Studies that use these devices have participants perform a cognitive task; the brain activity that is concurrent with the performance of the task is recorded. For example, a participant may be asked to form a visual image of a word that appears on a computer screen. The researchers can then determine which parts of the brain became active during imagery and in what order. Neuroscientists use other techniques as well. They study brain-damaged patients and the effects of lesions in laboratory animals, and use single- and multiple-cell recording techniques.

The Network Approach

The network approach is at least partially derived from neuroscience. In this perspective, mind is seen as a collection of individual computing units. These

units are connected to one another and mutually influence one other's activity via the connections. Although each of the units is believed to perform a relatively simple computation, for example, a neuron's either firing or not firing, the connectivity of the units can give rise to representational and computational complexity.

Chapter 7, which outlines the network approach, has two parts. The first involves the construction of artificial neural networks. Most artificial neural networks are computer software simulations that have been designed to mimic the way actual brain networks operate, or the functioning of neural cell populations. Artificial neural networks that can perform arithmetic, learn concepts, and read out loud now exist. A wide variety of network architectures have developed over the last thirty years.

The second part of the network chapter is more theoretical and focuses on knowledge representation—on how meaningful information may be mentally coded and processed. In semantic networks, nodes standing for concepts are connected to one another in such a way that activation of one node causes activation of other related nodes. Semantic networks have been constructed to explain how conceptual information in memory is organized and recalled. They are often used to predict and explain data obtained from experiments with human participants in cognitive psychology.

The Evolutionary Approach

The theory of natural selection proposed by Charles Darwin in 1859 revolutionized our way of thinking about biology. Natural selection holds that adaptive features enable the animals that possess them to survive and pass these features on to future generations. The environment in this view is seen as selecting from among a variety of traits those that serve a functional purpose.

The evolutionary approach can be considered in a quite general way and used to explain phenomena outside of biology. The field of evolutionary psychology applies selection theory to account for human mental processes. It attempts to elucidate the selection forces that acted on our ancestors and how those forces gave rise to the cognitive structures we now possess. Evolutionary psychologists also adopt a modular approach to mind. In this case, the modules correspond to "favored" cognitive capacities that were used by ancestors successful at solving certain problems. Evolutionary theories have been proposed to account for experimental results across a wide range of capacities, from categorization to memory, to logical and probabilistic reasoning, language, and cognitive differences between the sexes.

A variant on this theme is evolutionary computing, in which the rules of evolution are applied to create successful computer algorithms. An offshoot of this form of computing is artificial life. These are software simulations that mimic biological ecosystems. There is also neural Darwinism, which uses evolution to explain the formation of neural circuits. See Chapter 8 for more on these.

The Linguistic Approach

Linguistics is an area that focuses exclusively on the domain of language. It is concerned with all questions concerning language ability, such as: What is language? How do we acquire language? What parts of the brain underlie language use? As we have seen, language is a topic studied within other disciplines, for example, cognitive psychology and neuroscience. Because so many different researchers in different disciplines have taken on the problem of language, we consider it here as a separate discipline, united more by topic than by perspective or methodology.

Part of the difficulty in studying language is the fact that language itself is so complex. Much research has been devoted to understanding its nature. This work looks at the properties all languages share, the elements of language, and how those elements are used during communication. Other foci of linguistic investigation center on primate language use, language acquisition, deficits in language acquisition caused by early sensory deprivation or brain damage, the relationship between language and thought, and the development of speech recognition systems.

Linguistics, perhaps more than any other perspective discussed here, adopts a very eclectic methodological approach. Language researchers employ experiments and computer models, study brain-damaged patients, track how language ability changes during development, and compare diverse languages.

The Artificial Intelligence Approach

Researchers have been building devices that attempt to mimic human and animal function for many centuries. But it is only in the past few decades that computer scientists have seriously attempted to build devices that mimic complex thought processes. This area is now known as artificial intelligence (AI). Researchers in AI are concerned with getting computers to perform tasks that have heretofore required human intelligence. As such they construct programs to do the sorts of things that require complex reasoning on our part. AI programs have been developed that can diagnose medical disorders, use language, and play chess.

AI secondarily gives us insights into the function of human mental operations. Designing a computer program that can visually recognize an object often proves useful in understanding how we may perform the same task ourselves. An even more exciting outcome of AI research is that someday we may be able to create an artificial person who will possess all or many of the features that we consider uniquely human, such as consciousness, the ability to make decisions, and so on.

It is the development of computer algorithms and their testing, their comparison with empirical data or performance standards, and their subsequent modification that constitute the methodology of the AI perspective. Not all computer programs are alike, however. Researchers have employed a wide range of approaches. An early attempt at getting computers to reason involved the application of logical rules to propositional statements. Later on, expert systems, scripts, and fuzzy logic procedures, among others, were used. Chapters 10 and 11 give detailed descriptions of these techniques.

The Robotics Approach

Finally, we consider robotics. Robotics may be considered a familial relation to AI and has appeared on the scene as a formal discipline just as recently. Whereas AI workers build devices that "think," robotics researchers build machines that must also "act." Investigators in this field build autonomous or semi-autonomous mechanical devices that have been designed to perform a physical task in a real world environment. Examples of things that robots can do presently include navigating around a cluttered room, welding or manipulating parts on an assembly line, and defusing bombs.

The robotics approach has much to contribute to cognitive science and to theories of mind. Robots, like people and animals, must demonstrate successful goal-oriented behaviors under complex, changing, and uncertain environmental conditions. Robotics therefore helps us to think about the kinds of minds that underlie and produce such behaviors.

In Chapter 12 we outline different paradigms in robotics. Some of these approaches differ radically from one another. The hierarchical paradigm offers a "top down" perspective, according to which a robot is programmed with knowledge about the world. The robot then uses this model or internal representation to guide its actions. The reactive paradigm, on the other hand, is "bottom up." Robots that use this architecture respond in a simple way to environmental stimuli: they react reflexively to a stimulus input and there is little in the way of intervening knowledge.

In Depth: Categories of Mental Representation

We have said that there are three broad classes of mental representation—digital, analog, and propositional—each having its own characteristics, and we gave examples of each. However, the history of research in cognition suggests that there are also numerous forms of mental representation. Paul Thagard, in *Mind: Introduction to Cognitive Science* (2000), proposes four. These are concepts, propositions, rules, and analogies. Although some of these have already been alluded to and are described elsewhere in the book, they are central to many ideas in cognitive science. It is therefore useful to sketch out some of their major characteristics here.

A **concept** is perhaps the most basic form of mental representation. A concept is an idea that represents things we have grouped together. The concept "chair" does not refer to a specific chair, such as the one you are sitting in now, but is more general than that. It refers to all possible chairs no matter what their colors, sizes, and shapes. Concepts need not refer to concrete items. They can stand for abstract ideas, for example, "justice" or "love." Concepts can be related to one another in complex ways. They can be related in a hierarchical fashion, where a concept at one level of organization stands for all members of the class just below it. "Golden retrievers" belongs to the category of "dogs," which in turn belongs to the category of "animals." We discuss a hierarchical model of concept representation in the network approach chapter. The question of whether concepts are innate or learned is discussed in the philosophical approach chapter. The artificial intelligence chapter outlines the use of structures called frames as a means of representing conceptual knowledge.

A **proposition** is a statement or assertion typically posed in the form of a simple sentence. An essential feature of a proposition is that it can be proved true or false. For instance, the statement "The moon is made out of cheese" is grammatically correct and may represent a belief that some people hold, but it is a false statement. We can apply the rules of formal logic to propositions to determine the validity of those propositions. One logical inference is called a syllogism. A syllogism consists of three propositions. The first two are premises and the last is a conclusion. Take the following syllogism:

All men like football.

John is a man.

John likes football.

Obviously, the conclusion can be wrong if either of the two premises is wrong. If it is not true that all men like football, then it might not be true that John likes football, even if he is a man. If John is not a man, then he may or may not like football, assuming all men like it. Syllogistic reasoning of this sort is the same as deductive reasoning, mentioned earlier.

You may have noticed that propositions are representations that incorporate concepts. The proposition "All men like football" incorporates the concepts "men" and "football." Propositions are more sophisticated representations than concepts because they express relationships, sometimes very complex ones, between concepts. The rules of logic are best thought of as computational processes that can be applied to propositions in order to determine their validity. However, logical relations between propositions may themselves be considered a separate type of representation. The evolutionary approach chapter provides an interesting account of why logical reasoning, which is difficult for many people, is easier under certain circumstances.

Logic is not the only system for performing operations on propositions. Rules do this as well. A **production rule** is a conditional statement of the form: "If *x,* then *y,*" where *x* and *y* are propositions. The "if" part of the rule is called the condition. The "then" part is called the action. If the proposition that is contained in the condition (*x*) is true, then the action that is specified by the second proposition (*y*) should be carried out, according to the rule. The following rules help us drive our cars:

> If the light is red, then step on the brakes.
>
> If the light is green, then step on the accelerator.

Notice that, in the first rule, the two propositions are "the light is red" and "step on the brakes." We can also form more complex rules by linking propositions with "and" and "or" statements:

> If the light is red or the light is yellow, then step on the brakes.
>
> If the light is green and nobody is in the crosswalk, then step on the accelerator.

The "or" that links the two propositions in the first part of the rule specifies that if either proposition is true, the action should be carried out. If an "and" links these two propositions, the rule specifies that both must be true before the action can occur.

Rules bring up the question of what knowledge really is. We usually think of knowledge as factual. Indeed, a proposition such as "Candy is sweet," if validated, does provide factual information. The proposition is then an example of declarative knowledge. **Declarative knowledge** is used to represent facts. It tells us what is and is demonstrated by verbal communication. **Procedural knowledge**, on the other hand, represents skill. It tells us how to do something and is demonstrated by action. If we say that World War II was fought during the period 1939–1945, we have demonstrated a fact learned in history class. If we ski down a snowy mountain slope in the winter, we have demonstrated that we possess a specific skill. It is therefore very important that information-processing systems have some way of representing actions if they are to help an organism or machine to perform those actions. Rules are one way of representing procedural knowledge. We discuss two cognitive rule-based systems, the Atomic Components of Thought (ACT) and SOAR models, in the cognitive approach chapters.

Another specific type of mental representation is the **analogy**, although, as is pointed out below, the analogy can also be classified as a form of reasoning. Thinking analogically involves applying one's familiarity with an old situation to a new situation. Suppose you had never ridden on a train before, but had taken buses numerous times. You could use your understanding of bus riding to figure out how to take a ride on a train. Applying knowledge that you already possess and that is relevant to both scenarios would enable you to accomplish this. Based on prior experience, you would already know that you have to first determine the schedule, perhaps decide between express and local service, purchase a ticket, wait in line, board, stow your luggage, find a seat, and so on.

Analogies are a useful form of representation because they allow us to generalize our learning. Not every situation in life is entirely new. We can apply what we have already learned to similar situations without having to figure everything out all over again. Several models of analogical reasoning have been proposed (Forbus, Gentner & Law, 1995; Holyoak & Thagard, 1995). We outline some features of analogical reasoning in the Minds On section of this chapter. You can turn to this now if you want to try to solve an analogical reasoning problem. The application of the analogical approach in artificial intelligence is called case-based reasoning and is later described in the artificial intelligence chapter.

Minds On Exercise: Analogical Reasoning

To give you a sense of what it is like to reason analogically, we present here a classic problem, first posed by Duncker (1945). It is called "the tumor problem." Take a

moment to read it. After reading it, see if you can come up with the solution. If stumped, read past the next paragraph for the solution as well as an account of how the problem can be solved.

> Suppose you are a doctor faced with a patient who has a malignant tumor in his stomach. To operate on the patient is impossible, but unless the tumor is destroyed, the patient will die. A kind of ray, at a sufficiently high intensity, can destroy the tumor. Unfortunately, at this intensity the healthy tissue that the rays pass through on the way to the tumor will also be destroyed. At lower intensities the rays are harmless to healthy tissue, but will not affect the tumor. How can the rays be used to destroy the tumor without injuring the healthy tissue?

And here is another story. This one is called "the general and fortress problem." Please read it. Does it help you to come up with a solution to the tumor problem? How?

> A small country was ruled from a strong fortress by a dictator. The fortress was situated in the middle of the country, surrounded by farms and villages. Many roads led to the fortress through the countryside. A rebel general vowed to capture the fortress. The general knew that an attack by his entire army would capture the fortress. He gathered his army at the head of one of the roads, ready to launch a full-scale direct attack. However, the general then learned that the dictator had planted mines on each of the roads. The mines were set so that small bodies of men could pass over them safely, since the dictator needed to move his own troops and workers to and from the fortress. However, any large force would detonate the mines. Not only would this blow up the road, but it would also destroy many neighboring villages. It seemed impossible to capture the fortress. However, the general devised a simple plan. He divided his army into small groups and dispatched each group to the head of a different road. When all was ready, he gave the signal and each group marched down a different road. Each group continued down its road to the fortress, so that the entire army arrived together at the fortress at the same time. In this way, the general captured the fortress and overthrew the dictator.

You may have noticed a number of similarities between these two stories. The tumor is similar to the fortress. The rays that were to be used to destroy the tumor are like the soldiers sent to capture the fortress. The healthy tissue in the first story can be likened to the villages in the second. Noticing these similarities, you may have then applied a solution similar that used by the rebel general to the problem of eradicating the tumor. Like the solution of dividing up the army and sending its

soldiers down separate roads to converge on the fortress, the solution to the tumor problem involves dividing up the high-intensity ray into multiple low-intensity rays and then targeting them on the tumor at different angles. In this fashion, the rays converge on the tumor in a show of strength and destroy it without damaging any of the surrounding healthy tissue. Gick and Holyoak (1980) found that only 10% of participants in their study could solve the tumor problem correctly in the absence of their being provided with the general and fortress story. A full 75% of participants solved it correctly when they were provided with the story.

Models of analogical reasoning usually posit a novel analog that stands for a new situation or problem to be solved. That is the tumor problem in this example. They also employ an existing analog that has been derived from another learned situation. That is the general and fortress problem. The analogy is the systematic relationship between these two analogs and includes the similarities pointed out above. Most models of analogical reasoning reveal that there are four stages to the process of analogical reasoning. First is comprehension of the target problem. Second is remembering a similar source problem for which a solution is already known. Next, the source and target problems are compared. This is equivalent to mapping out similarities in their corresponding structures. As a last step, the source problem is adapted to produce a solution to the target problem.

Food for Thought: Discussion Questions

1. Many metaphors have been proposed for thinking about the mind. These range from water pumps to telephone systems. Can a corporate office building serve as a metaphor for mind? Why or why not?
2. Might concrete concepts such as "snake" be represented differently from abstract concepts such as "democracy"? Which kind of concept lends itself more easily to an analog representation? Why?
3. Describe how a handheld pocket calculator performs division at computational, algorithmic, and implementational levels of analysis.
4. Images, concepts, propositions, rules, and analogies are all forms of mental representation. Can you think of other examples?
5. Think of an instance in everyday life in which you used analogical reasoning. Describe in as much detail as possible the target and source problems, and the similarities between them.

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Franklin, S. (1995). *Artificial minds*. Cambridge, MA: MIT Press.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart & Winston.

Sobel, C. P. (2001). *The cognitive sciences: An interdisciplinary approach*. Mountain View, CA: Mayfield.

Thagard, P. (2000). *Mind: Introduction to cognitive science*. Cambridge, MA: MIT Press.

2

The Philosophical Approach: Enduring Questions

"What is Matter?—Never mind."
"What is Mind?—No matter."

—Anonymous, 1855

What Is Philosophy?

Philosophy in its broadest sense is the search for wisdom and knowledge. It is the first approach we will tackle in our voyage through the different disciplines of cognitive science. There are good reasons for beginning here. Philosophy plays a vital participatory role in cognitive science. It does this not by generating results, since it is a theoretical rather than experimental discipline, but by "defining problems, criticizing models, and suggesting avenues for future research" (Garfield, 1995, p. 374). More than any other discipline in cognitive science, philosophy is not limited by its subject matter or a particular theoretical stance. It is therefore free to evaluate and contribute to the remaining disciplines in a way the others cannot. This approach is also the oldest of the different approaches, tracing its origins back to the ancient Greeks. It is thus fitting that we begin our tour here.

The translation of the word philosophy yields "love of wisdom," indicating the philosopher's concern with knowledge and with understanding the universe. Philosophy as a formal discipline studies a wide range of topics. In fact, there is no topic that is not fair game for a philosopher; he or she may examine politics, ethics, esthetics, and other subjects. We concern ourselves here with two branches of philosophy. **Metaphysics** examines the nature of reality. The mind-body problem is a metaphysical one at heart, because it seeks to understand whether the mental world is part of the physical material world. **Epistemology** is the study of knowledge and asks such questions as: What is knowledge? How is knowledge represented in the mind? How do we come to acquire knowledge?

In this chapter we will survey philosophic thoughts that center on four vexing issues, most of which are summed up in terms of "this" versus "that." This terminology suggests that the debates that have arisen from these issues have polarized the arguments and that there are only two possible answers to a problem. We will see that this is actually not the case and that there are multiple ways to conceptualize the issues. These issues are the mind-body, free will-determinism, and nature-nurture debates. In addition, we discuss the question of consciousness and its relation to cognitive science.

The Mind-Body Problem

The mind-body problem addresses how psychological or mental properties are related to physical properties. The debate stems from a fundamental conception about what the mind is. On the one hand we have the brain that is material and physical. It is made up of substances that we can measure and understand. The mind could be thought of in the same way, as simply a physical thing. On the other hand, there are those who argue that the mind is something more. They say we can't equate our subjective conscious experiences, such as beliefs, desires, and thoughts, with something as mundane as the brain. They say the mind is nonphysical and consists of something resembling a soul or spirit. The mind as a nonphysical entity inhabiting the brain or other physical entity is sometimes called "the ghost in the machine."

The first question of the mind-body problem refers to the nature of what mind is. Is the mind physical or something else? A second and more specific question concerns the relationship between these two entities. If we assume that there are two such entities, then what is the causal relationship between them? Does the mind control the mind or does the body control the mind? Table 2.1 shows the possible relationships between mind and body and the labels that go with each.

Table 2.1 Different interpretations of the mind-body debate and the schools of thought associated with each

Class of Theory	*Name of Theory*	*Physical Universe*	*Causal Direction*	*Mental Universe*
Monism				
	Idealism/Solipsism	None	No causality	Mind
	Physicalism	Body	No causality	None
Dualism				
	Classical Dualism	Body	←	Mind
	Parallelism	Body	No causality	Mind
	Epiphenomenalism	Body	→	Mind
	Interactionism	Body	↔	Mind

Source: Adapted from Kitzis, S. N. (2002). *Mind and meaning: A flight of imagination, a voyage of discovery*. Pacific Grove, CA: Wadsworth Thompson.

Our discussion in this section will be structured around basic conceptions of the nature of mind. According to **monism**, there is only one kind of state or substance in the universe. The ancient Greek philosopher Aristotle (384–322 B.C.E.) was a monist (Figure 2.1). He characterized the difference between mind and body as the difference between form and matter. One way to think of his notion is to consider a lump of clay. It is made up of physical matter and we can think of it as corresponding to the brain. We can shape the clay with our hands into different forms—for example, we can roll it into a ball or flatten it out into a pancake. The shapes the clay can assume, Aristotle implied, are like the different thoughts the mind can take on when it undergoes different patterns of activity. These shapes are just different physical states and do not constitute any nonphysical or spiritual substance. There are two classes of monist. Monists who believe only in mental substance are either idealists or solipsists. Monists who believe exclusively in physical substance, like Aristotle, are physicalists.

In **dualism**, one believes that both mental and physical substances are possible. Plato, another Greek philosopher (427–347 B.C.E.), was a dualist. Plato was Aristotle's teacher, but the two held quite different views. Plato believed that the mind and the body exist in two separate worlds. Knowledge of the mind, he thought, exists in an ideal world of forms, which is immaterial, non-extended, and eternal. The body resides in a world that is material, extended,

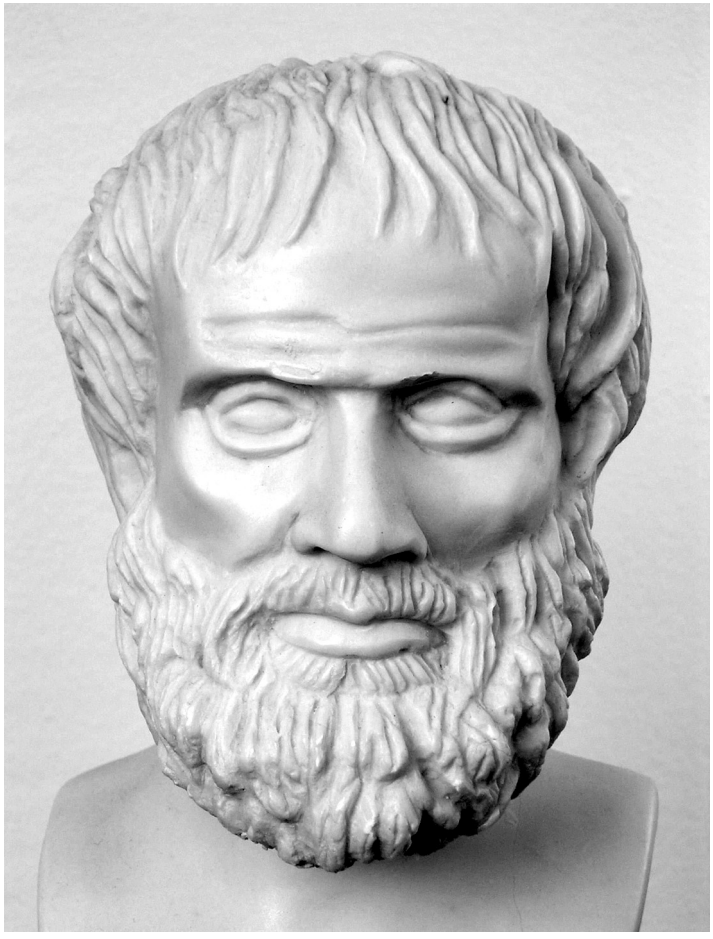


Figure 2.1 A bust of the early Greek philosopher Aristotle. He believed there was no substantial difference between mind and matter

and perishable. There are crucial differences between the objects of one world versus those of the other. Mental ideas such as "circle" that reside in the ideal world of forms are perfect, according to Plato: the circles of this world are always perfectly round. Concrete examples of circles that we find in the real world are always imperfect. If we examine an actual circle, at some level of magnification the circle's edge will lose its roundness. There are a variety of different modern dualist schools, each based on a particular view of the relationship between the mental and the physical. These include classic dualism, parallelism, epiphenomenalism, interactionism, and functionalism.

Flavors of Monism

Idealism has been attributed to the Irish philosopher George Berkeley (1685–1783). In this philosophy, there is only the mental realm. There is no physical realm. Our brains, bodies, and everything else in the universe exist only as concepts in God's mind. All our experiences are the same as God's experiences and we have them because God has them. Problems with the idealist position should become immediately clear. It resorts to a religious and mystical form of explanation and must be accepted on faith. Although it is logically consistent, this view cannot be tested and so is not considered scientific.

Solipsism also falls into the category of "mental only" theories of mind. According to this perspective, the universe exists only in one's mind. In other words, each person, by virtue of having a mind, creates his or her own world. This brings up a host of questions. If the universe exists only in each individual mind, then there must be as many universes as there are individuals. If this were the case, then which universe is the right one? Do they all exist at the same time? This view is inherently subjective and, like idealism, is not subject to experimental scrutiny.

On the flip side of our metaphysical coin is **physicalism** or materialism. The origins of this view go back to the Greek philosopher Democritus (ca. 460–370 B.C.), who believed all things were composed of atoms. The attributes and behaviors of the atoms, he said, can explain the differences between things, including the differences between mind and body. Physicalists, like idealists and solipsists, are also monistic, and believe that the universe is composed of a single substance. They however regard this substance as physical and material rather than spiritual or ethereal. Physicalism is thus the doctrine that everything that exists is physical. The operations of the mind are seen here as simply the operations of the brain.

A school of physicalism known as **reductive physicalism** is an example of **reductionism**, where one theory or view is used to completely account for another. As an example, it is often believed that neuroscience, which is concerned with the study of the brain's physical make up and processes, will ultimately be able to account for all psychological or mentalistic levels of description. To illustrate, a psychologist's explanation of an anxiety disorder that uses a mentalistic term such as "fear" may someday—this class of physicalists hopes—be reduced to a description of neurotransmitter changes in brain structures, such as the amygdala. A second breed of physicalists go by the name of **nonreductive physicalists**. They believe physical processes can give rise to emergent and irreducible mental phenomena. This school believes that one cannot completely do away with mentalistic description. We will discuss the property of emergence in the context of consciousness.

Physicalism has received its share of critique as well. Some allow that physical processes can determine mental ones but deny that they can explain them. So, they argue, changes in the amygdala may very well correlate with and produce fear, but they do not explain different kinds of fear, how a person becomes fearful, and so on. These critics acknowledge the world is physical but indicate that for many phenomena there is no physical explanation. In these cases, they believe, it is perhaps better to explain using mental terms.

Flavors of Dualism

Now that we have reviewed the various forms of monism, let us turn our attention to its logical alternative. Dualists believe both mental and physical realms are possible, but differ in the way they think these two interact. **Classical dualism** originated with the French philosopher René Descartes (1596–1650). Descartes was a revolutionary philosopher for his time and introduced theories on many of the ideas that underlie cognitive science. He believed in a one-way causal link, with the mind controlling the body but not vice versa. Descartes thought the mind exerted its control on the body through the pineal gland, perhaps because it is one of the few anatomical structures not duplicated on either side of the brain (see Figure 2.2). In this view, the mind is like a puppet master, the body is like a puppet, and the pineal gland is like the puppet strings, by which the former controls the latter. Classical dualism conforms to most people's common-sense notion of the mind-brain relationship, which is that our thoughts control our actions. For instance, when we feel hungry, we get up and eat a snack. It certainly seems as if the feeling of hunger comes first and causes the action of getting up to eat.

There are three other dualist schools of thought. In **parallelism**, the mind and body are distinct and isolated from each other. One cannot have an effect on the other. But there certainly appears to be some sort of causal link between them. Otherwise, how is it that we can move our legs when we decide to walk? A parallelist response would be that God or some other unknown force has synchronized the workings of the mind and body so that they work seamlessly together. It is like having two clocks, a physical and a mental one, that were both started at the same time and run alongside one another, keeping the same time. The critique here is the same as for idealism. This view resorts to mystical explanation and so is not accepted within mainstream cognitive science.

The **epiphenomenalist** school allows the physical to cause the mental, but prohibits causation in the other direction. In this account, the mind is like a side effect of the brain's operations, but cannot in any way feed backward to

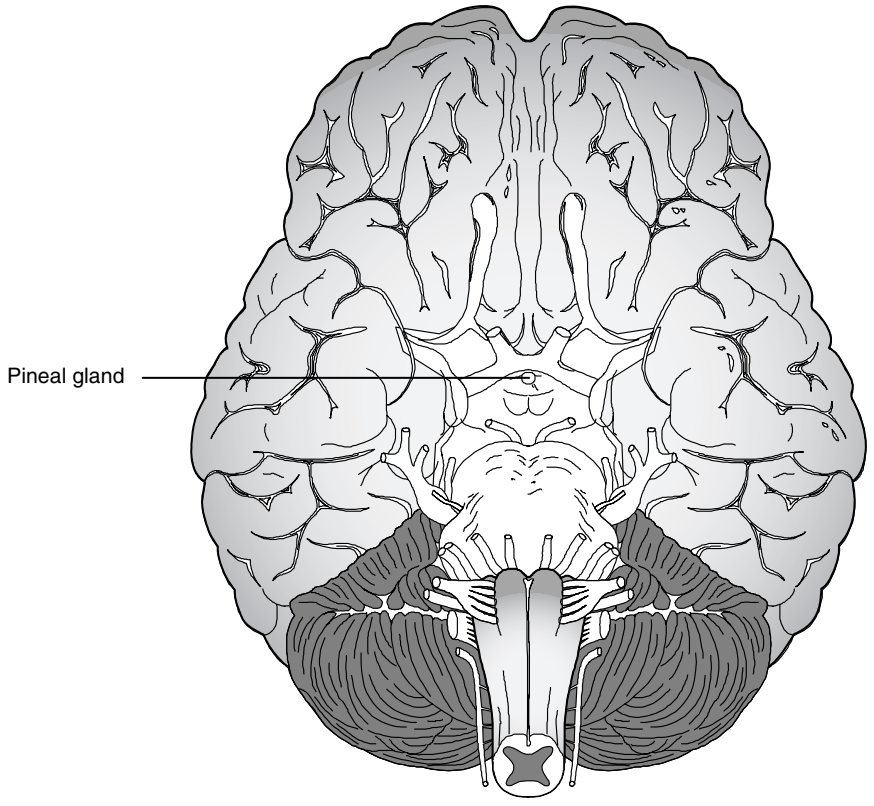


Figure 2.2 Descartes believed the pineal gland to be the location where the mind influenced the body. This belief was itself influenced by the fact that the pineal gland is located in the center of the brain

influence the brain. As an analogy, imagine the exhaust coming out of a car engine. The engine is the brain, while the exhaust it creates would be the mind. The engine produces the exhaust, but the exhaust in no way affects the operation of the engine or car. A difficulty with this view is that it runs counter to our introspections, many of which place thoughts before actions.

Interactionism allows causality to travel both ways. The body can affect the mind and the mind can also affect the body. Here we have a two-way street that allows each realm to influence the other. One can have a thought that produces an action, but also a physical brain activity that produces a thought. The approach has its problems: how these two mutually influence one another as an integrated whole is not specified. Of particular concern is how the mental can affect the physical. It is easier to conceptualize how physical causes have

effects. We already have a good account of this in Newtonian physics. Immaterial causation is more of a mystery.

Evaluating the Dualist Perspective

One critique of dualism comes from the philosopher Gilbert Ryle. Ryle's argument (1949) centers on our conceptualization of mind and its relation to the body. He believes that the mind is not any particular component of the brain, but all the parts working together as a coordinated, organized whole. He illustrates with a story. Imagine a visitor from a foreign country arriving at a large university. He is shown around the campus and the various parts of the school are pointed out to him, including the dormitories, departments, and lawns. The visitor, who has never seen any of this before, is puzzled. He says; "Well, I've seen all of this, but I haven't yet seen the university." We would have to explain to him that the university is not any of the individual sites he has viewed, but all the sites together and the interconnections among them (see Figure 2.3). Ryle thinks philosophers fall into the same trap as the visitor, mistaking the part or parts for the whole. He argues that the mind belongs in a conceptual category different from that of the body, just as the university is in a category different from those of the things that make it up.

Andy Clark (2001) summarizes several other critiques of dualism. These would apply to Descartes' conception as well as other perspectives. Clark says that dualism is uninformative and tells us what the mind isn't, rather than what it is. If the mind isn't the brain and isn't physical, then what is it? Dualists are remarkably silent on this matter, often conceding that it is something nonphysical that we can't understand yet. As a theory, dualism is also inelegant, because it postulates two worlds that must be coordinated. An explanation that does not violate the principle of Occam's razor (that the simpler explanation is usually the correct one) would involve a single type of world, not requiring coordination.

There are further problems with dualism. One has to do with the dependence of the mental on the physical. Factors that affect the brain such as head trauma or drug use have direct and dramatic mental effects. We can see that damage to a certain part of the brain, say, from a motorcycle accident, results in specific forms of mental disruption, for example, language deficits. Taking a drug like marijuana, which alters brain chemistry, results in altered mental states. In addition, the evolutionary approach shows us there is a general positive correlation between brain size and intelligence across species, with larger

Figure 2.3 Where is the university?

Source: Courtesy of Peter Finger Photography 2003.

brain sizes linked to increased cognitive capacity. It is obvious from these observations that the mental is integrated with the physical, that the mind depends on the brain.

Some dualists in response to attacks on their positions have stated that the mind exhibits extraordinary abilities and that it is or will be impossible for a physical system to duplicate such abilities. For instance, how can a physical system, be it a brain or a computer, write a novel or negotiate a peace treaty? The truth is that as our technological sophistication increases, many of these abilities are becoming better understood and implemented computationally. There are now computers that can beat the best chess champions and successfully diagnose medical disorders. These are capacities once thought to be the exclusive domain of humans.

Dualists and other philosophers also argue that our subjective experiences—things like thoughts, beliefs, and desires—are not equivalent to physical brain

states. They base this conclusion primarily on introspection. When we examine what is inside our heads, they say, these subjective experiences seem to be something more than just physical. The problem with this argument is that introspection is a weak form of evidence and can be wrong (as are many of our ideas). What is required is objective proof that such experiential states are not physical.

Functionalism

The most influential philosophical theory of mind in cognitive science is functionalism. For this reason we will discuss it in considerably more detail than any of the theories we've already discussed. To get an idea of what functionalism is about, we need to make a distinction between two ways of classifying things. **Physical kinds** are identified by their material composition only. In this view, jellyfish and carpets are different because they are made up of fundamentally different physical substances. **Functional kinds** however are distinguished by their actions or tendencies. Here, we could say that all automobiles fall under the same functional category because they do the same things, namely, transport goods and people, even though they may be made up of different elements.

So far so good, but things get interesting when we extend these ways of classifying to the idea of mind. If we think of mind as a physical kind, then minds must be the same things as brains, since, as far as we know, minds cannot exist apart from physical brains. To many, this seems too exclusive. It is possible, they argue, that computers might develop minds and that there might be alien species with minds (see Figure 2.4). Neither computers nor aliens need have brains in the sense that we know them. It is more fruitful, they say, to identify minds as functional kinds and to define them by the sorts of processes they carry out rather than the stuff they're made of. According to **functionalism**, mental states are not just physical states, but also the functioning or operation of those physical states. According to this view, a mind could conceivably be implemented in any physical system, artificial or natural, capable of supporting the appropriate computation.

Functionalism has several significant implications (Garfield, 1995). One is that the same mental state could be realized in a quite different way in two separate physical systems. This can be illustrated with computing devices. Two such different devices, say a desktop computer and a palm-sized personal data assistant, can both compute the same result, such as displaying a page of text, but in entirely different ways. The same might also be true for human

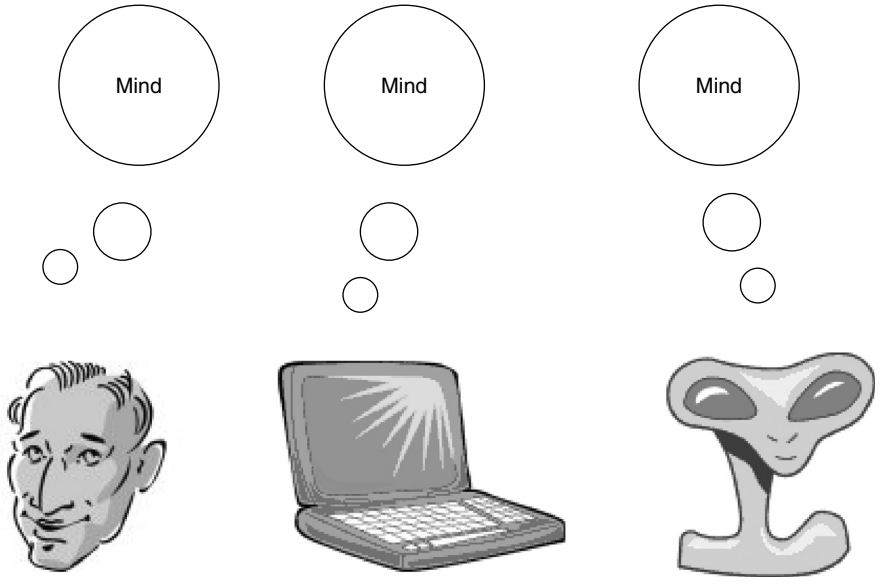


Figure 2.4 According to functionalism, different physical substrates can in principle all give rise to mind

computation. If we examined the brains of two people thinking exactly the same thought, we would in all likelihood not find exactly the same processes at work.

There are several schools of thought in functionalism. These range from conservative views that advocate direct connections between physical and computational states, to more liberal ones that emphasize computation over physicality. The liberal schools give two reasons for their stance. They say that for both computers and thinking organisms, the number of possible computational states always exceeds the number of possible physical states. Take for example all the different possible beliefs one could hold concerning politics, the environment, one's friends, and so on. Mathematically, the number of such beliefs is infinite (Garfield, 1995). The number of possible physical states the brain can assume, though, is finite. A computational level of description thus becomes a richer and more diverse way of describing the mind and should be the preferred level. Second, liberal functionalists argue that psychological states such as beliefs are defined more by their relations to other such states, to inputs from the environment, and to behaviors than their relations to physical states. A belief such as "patriotism" usually manifests itself in other beliefs, for

example, in flag-waving. It will elicit predictable reactions to environmental stimuli, for example, feeling upset when one's country has been criticized, and will produce external behaviors such as marching or protesting.

To summarize, functionalism implies that mental states might not be reduced to any particular physical state. This argument does not require us to be dualists. It is not saying that mental states don't conform to physical ones, only that there may be a wide variety of possible physical states capable of producing any given mental state.

Evaluating the Functionalist Perspective

Although functionalism has been the dominant view in cognitive science since the 1970s, it is not without its deficiencies (Maloney, 1999). Remember that a tenet of functionalism is that minds that are not based on brains can exist. They can exist in things like computers as long as the physical substrates of those objects allow for the relevant computations. Critics have argued that, although it is possible that minds can exist in the absence of brains, this does not make it plausible. There is no current empirical evidence to justify this claim. We have yet to see something mental in the absence of a brain. Also, some have argued that the failure to identify mind with a physical kind can itself be considered reason to do away with the concept of mind—rather than give it special status as a functional kind.

An additional problem with functionalism is that it cannot account for the felt or experienced character of mental states—a phenomenon known as **qualia** (quale, singular). Examples of qualia include the subjective experience of what it is like to feel "hungry," to be "angry," or to see the color "red." It would seem that these kinds of experiences cannot be replicated as purely functional processes. A machine could be programmed to "see" the color red, even mimicking the same human functional process, but this machine could not have the same experience of what it is like to see red that a person has.

What is more, two individuals having the same conscious experience often do not experience it subjectively in the same way. A number of experiments have shown this to be the case with color perception. Participants looking at the same color will describe it differently (Chapanis, 1965). If asked to point out on a color spectrum what pure green looks like, one person may select a yellow-green, another a blue-green. This is the case even though the functional operations of their respective brains as they view the color are approximately equivalent. In this case, the neurophysiological operations behind color perception tend to be the same across individuals.

The Free Will–Determinism Debate

The Issue of Determinism

Think for a minute about all the decisions you've made in your life. Some of these were important, for example, deciding which college or university to attend, or perhaps deciding whether or not to pursue a romantic relationship. Others may have been less essential, for example, whether to get pork with fried rice or Szechuan chicken for lunch. The free will–determinism debate is about whether these behaviors are within our control. Did you consciously ponder the consequences of attending one school versus another, thinking of the possible pros and cons of each, or did forces beyond your control push you to attend one school over the others? Those who take the free will side of this debate argue that individuals independently initiate their own actions. Those favoring determinism argue that actions can be explained in terms of the initiating causes that precede them, which implies that individuals act only dependently, as a consequence of these causes.

Let us start with **determinism**. It is the view that all physical events are caused or determined by the sum total of all prior events. Our actions, which are physical events, must therefore also be determined. Suppose you get up to grab a snack from the refrigerator. According to determinism, your getting up was caused by another physical event (or events) that immediately preceded it. This event may have been a stomach pang or hearing a restaurant advertisement on the radio. Furthermore, the determinist view is that it was inevitable that you would have undertaken this action because of the preceding event. In other words, there was no other action you could have taken, given the nature of who you are and the set of events that preceded the act of getting up.

The philosopher David Hume (1711–1776) suggests we conceive of a deterministic universe in terms of billiard balls. To get a sense of this notion, imagine that, initially, a set of billiard balls is scattered across a table and the balls occupy random positions. We then come along with a cue stick and knock one ball into another. Each moving ball, in this Humean model of causation, is an event that was caused by a preceding event and in turn causes another (see Figure 2.5).

This model has several implications. The first, as mentioned before, is determinism. If ball A strikes ball B, then there is only one "action" that ball B can take, that action being entirely caused by ball A's action. Of course, this is a simplified model. It is possible to have multiple forces acting on ball B, in which case their conjoint influence determines ball B's action. The point is that ball B's action is determined completely by the forces acting on it in the

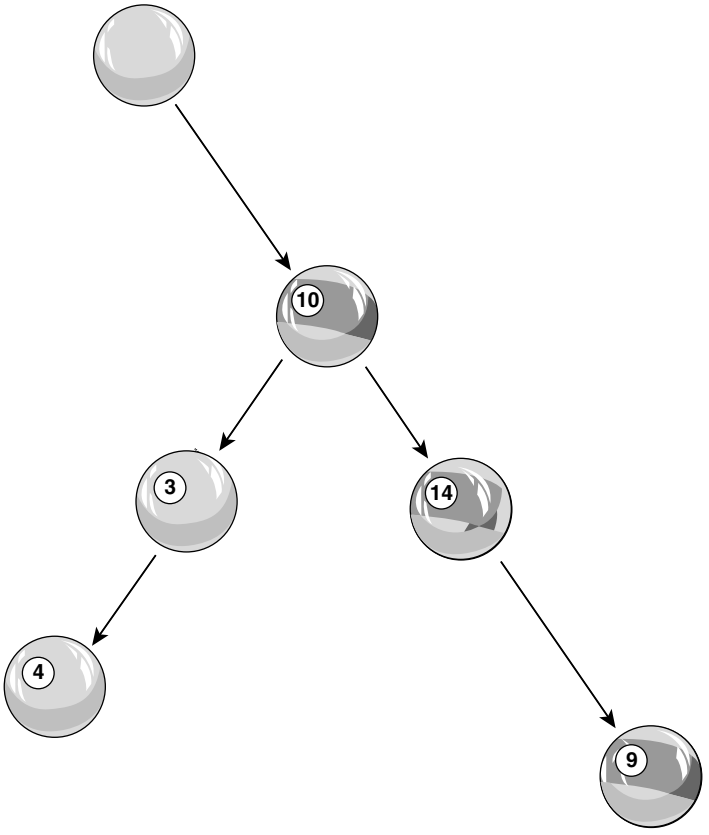


Figure 2.5 In the Humean model of causation, action in the physical universe can be likened to billiard balls on a billiard ball table, where balls bump into one another. In this instance the cue ball initiates two causal paths, one involving balls having a stripe, the other solid-colored balls

temporally preceding moment. A second implication is **replication**. If we were to return all of the balls to their original positions and strike ball A with the cue stick again, in exactly the same way, then all of the balls would move once more in exactly the same way. The third implication is **prediction**. If we know the positions of all of the balls and the conditions under which the first ball will be hit, such as the force to be applied and angle of the cue stick, then we can know ahead of time what all of the balls on the table will do. Many of these properties of a causal system also happen to underlie scientists' understanding of the way that variables in a controlled scientific study interact. The billiard balls' actions on each other describe the way experimental variables are

hypothesized to influence each other. Indeed, replicability and predictability are two of the cornerstones of the scientific method. We will talk more about the scientific method in the psychology chapter.

If we translate this deterministic model of causality to human behavior, it means we can also predict any action a person might undertake for the rest of his or her life. All that is needed is an understanding of the system and the forces acting on it. In this case the system is the brain. The forces that act on the brain can include influences such as a stomach pang or radio advertisement. The behavior of the system would be the overt behavior the brain gives rise to, such as getting up to get a snack. This notion is not too far removed from what the behaviorist psychologists in the early 20th century were proposing. In their case they didn't purport to understand the mind or how it operated; that is, they ignored the system itself. They instead concentrated on understanding the causal relationship between stimulus inputs from the environment and the consequent behavioral output. Some behaviorists felt they could entirely predict a person's actions on the basis of that person's conditioning history, which is the schedule of rewards and punishments he or she has undergone so far in his or her life. We will talk more about behaviorism in the following chapter.

The Issue of Free Will

Most people find the Humean billiard ball model of causality unsettling. This is because it turns us into automatons reacting in known ways to the forces impinging on us. We would like to believe that we instead choose our own course of action. The free will perspective regards behavior as stemming from a decision or act of will. The decision is made autonomously and is itself not under the influence of any preceding causal factors. In other words, the will of the individual is the sole determining cause of the action. The will itself is not the product of any other cause, and is considered to be its own cause. People, according to this view, are thus not at the whim of forces beyond their control, but have the power to independently initiate their own actions.

The 20th century philosopher Ayn Rand (1963) has formulated an **entity model of causation** that, according to Rand, underlies free will. In this model, entities with specific identities are the cause of actions. An entity is defined roughly as an object capable of independent action. A person is an example of an entity. The actions an entity undertakes are determined not by some antecedent factor that acts on the entity, but instead by the nature of that entity. If an entity is a certain way, then it will act only in accordance with that way.

In fact, it is not possible for an entity to act in a way that is contradictory to its nature. By making entities rather than actions the cause of actions, Rand shifts the force of causality away from the environment and toward individuals.

Rand further argues that humans are beings of volitional consciousness. By this she means that although people are capable of thinking, they must make the decision to do so. By this account, both thought and the decision to think or not are part of human nature. If we decide to think, then we consciously control our actions. If we fail to think, then we are at the whim of our mind's subconscious associational processes. This notion of voluntarily directing one's thought processes Rand calls focus. Cognitive psychologists refer to it more broadly as attention. We discuss several models of attention in the cognitive approach chapter.

Evaluating the Free Will–Determinism Debate

A problem with free will is that it violates one key assumption of causality, which is that all events must have a cause. In this assumption, actions can't cause themselves. The causal universe is seen here as a network of dependencies, with actions existing only as possibilities until they are actualized by a triggering event or events. If this is true, then how can a person initiate an action "out of nowhere"? The strict version of free will implies that the decision or will of an individual to act is a cause that is itself uncaused. This cannot be accounted for in the scientific and causal view of the universe.

According to the cognitive science view, a decision is itself a mental process. Like other such processes, it should have informational inputs, perform a computation, and have corresponding informational or behavioral outputs. The decision to decide which school to attend must have as a starting point a mental list of possible schools along with data about each. The information, in this view, is then passed through some sort of decision-making algorithm. This algorithm might calculate a winning school by adding up scores pertaining to certain features such as prestige and geographic location. The output of the computation would be the single school to which one would then apply.

Using the above model, the decision is not uncaused, at least in the sense of having no preceding events. These events are the list of schools and farther back in time the decision to go to school in the first place. From a scientific and causal perspective, if the decision-making algorithm and the data fed to it are completely known, then the decision in principle should follow deterministic consequences, that is, the decision should be determined, replicable, and predictable. However, empirical evidence in the cognitive sciences rarely produces models that adhere perfectly to all three of these criteria. Part of the reason is

incomplete knowledge. If the algorithm or data is only partially understood, then the decision can be only partially determined, replicated, or predicted.

One way to reconcile the free will–determinism dilemma is to allow both to be true. This is what **compatibilism** offers. Compatibilists allow that free will and determinism can be reconciled and/or compatible with each other. It allows these two schools of thought to coexist. Compatibilists believe we have free will in the broad sense of moral responsibility, meaning that even though we may be constrained to act in a certain way, we always have the freedom to choose otherwise. For instance, imagine a thief pointing a gun at your head and asking for your wallet. You are certainly constrained under these circumstances to fork over your cash. But in a compatibilist world you don't have to do so, you are perfectly free to refuse and accept the consequences. Just as compatibilists do not deny causality, they believe every event has a cause and that our actions are always preceded by a cause. However, these antecedent events do not fix our actions. In any past circumstance, we would have been free to embark on some other alternate course of action.

In contrast to this is the theory of **incompatibilism**. Incompatibilists see free will and determinism as irreconcilable, meaning they are not both simultaneously true. This school argues that a person cannot be truly free in the moral sense if preceding causal events impact on his or her actions. Causal laws are seen here as preventing us from being true "free agents," from having absolute moral responsibility for our actions. One group of incompatibilists, called **libertarians**, believes we do have free will, that free will is not compatible with determinism, and that determinism must therefore be false. The problem with this view is that its adherents must show how we can be morally responsible in an indeterminist world (one where determinism is false). A second school, whose adherents are called **pessimists** or no-freedom theorists, argue that moral free will is impossible to prove. The pessimists say there are circumstances in which we can be free agents—but those circumstances occur only when we are unconstrained, when circumstances allow us to choose to do what we want. This however is not enough to explain free will in the larger sense, which is that we always have moral responsibility for our actions, regardless of the constraining factors.

The Knowledge Acquisition Problem

A fundamental question asked by even the earliest of philosophers was: How do we acquire knowledge? Clearly, you are not born knowing everything, otherwise you would not need to go to school, or wouldn't be reading this book.

But are we born knowing anything at all? Is the mind completely blank or do we start with some rudimentary understanding of the world? One way to frame these questions is within the **nature-nurture** debate. This debate centers on the relative contributions of biology and experience in determining any particular capacity. The term *nature,* in this context, refers to traits that are genetically or biologically determined. These are coded for in our genes and so are "hardwired," meaning they are present at birth or appear at a specific time during development. The term *nurture* refers to traits that are learned through experience and interaction with the environment. We will examine theories of knowledge acquisition that argue for the greater influence of one or the other.

According to **nativism,** a significant body of knowledge is innate or "built into" an organism. In this sense, nativism is a theory of knowledge that favors nature over nurture. Plato was the first to outline a nativist theory of knowledge. He thought learning was a matter of recollecting what is already known—these concepts existing in the ideal world of forms and being part of our immortal soul. **Rationalism** must be subtly distinguished from nativism. Descartes was the progenitor of this perspective. Rationalists also believe in the existence of innate ideas. These basic concepts include such ideas as "God" and "triangle." However, they additionally emphasize the existence of innate reasoning powers. These include certain logical propositions, such as knowing that something cannot exist and not exist at the same time. We can use these a priori rational powers to form new ideas that are not given to us innately. Descartes would agree that we are not born with the idea of "table," but can acquire it given our innate ability to perceive and think about objects.

Empiricism alternatively sees knowledge as acquired through experience: it favors nurture over nature. In this view, knowledge gets into the head through interaction with an environment, meaning it is learned. The senses provide the primary channels via which knowledge of the world is born. Our knowledge of the concept "lemon" in this account begins with looking at a lemon, touching and tasting it. The British philosopher John Locke (1632–1704) is credited as the founder of the empiricist movement. He used the phrase *tabula rasa,* which literally translates as "blank slate." Locke believed that we are born as blank slates, lacking any knowledge, and that over time experience puts writing onto the slate, filling it up.

Locke had a more fully developed theory of learning. He differentiates between simple ideas and complex ideas. **Simple ideas** are derived through sensory input or simple processes of reflection. They are received passively by the mind and cannot be reduced to simpler ideas. Looking at a cherry would generate the simple idea of "red." Tasting a cherry would produce the simple idea of "sweet." **Complex ideas** are formed from the active mental combination of simple ideas. They are created through reflection only and can be reduced to

parts, their component simple ideas. The idea of "cherry" would result from the associative combination of such simple ideas as "red," "sweet," and other commonly occurring sensations derived from one's experiencing cherries. This cluster of simple ideas is naturally associated because each time we experience a cherry it comes to mind. For this reason, Locke and others who have proposed similar notions are sometimes known as the associationists.

Evaluating the Knowledge Acquistion Debate

One might be tempted to immediately dismiss the doctrine of innate ideas put forth by the nativists and rationalists. After all, it seems absurd that we should be born knowing factual information such as the content of the Gettysburg address. But the scope of knowledge is broader than this. Think back to the previous chapter, in which we defined declarative knowledge for facts and procedural knowledge for skills. There is quite a bit of research supporting the notion that some forms of procedural knowledge are innate. Newborn infants for instance come into this world with a variety of different skills. These skills are universal across the human species and manifest themselves so soon after birth that they couldn't possibly have been learned. They therefore qualify as examples of innate knowledge. Let us examine a few of these.

All infants demonstrate a set of **reflexes**. These reflexes include the grasping reflex, in which the fingers tighten around a touch to the palm, and the rooting reflex, in which the infant turns his or her head and begins sucking an object placed near the mouth. Reflexes serve a clear adaptive function. Grasping and sucking, along with behaviors generated by other early reflexes, are important for survival. The physiology behind reflexes is simple and fairly well understood. A stimulus triggers one or more sensory neurons that then activate intermediary neurons. These in turn activate motor neurons, causing the resulting behavior. It is easy to see how such a simple mechanism could be hardwired at birth to enable the infant to respond effectively to its environment. Figure 2.6 shows the anatomy of a spinal reflex.

Smell preference is another example of innate behavior. Steiner (1979) found newborns tend to agree with adults in terms of which odors they consider pleasant or unpleasant. He found that odors such as strawberry and banana elicited agreeable facial expressions from young infants, for example, smiling. Unpleasant odors such as fish and rotten eggs elicited expressions of disgust. As is the case with reflexes, these smell preferences have survival value. Babies who find the smell of fruit attractive will eat the fruit and thereby gain nutrients, those that are repulsed by spoiled or unpalatable food will reject the

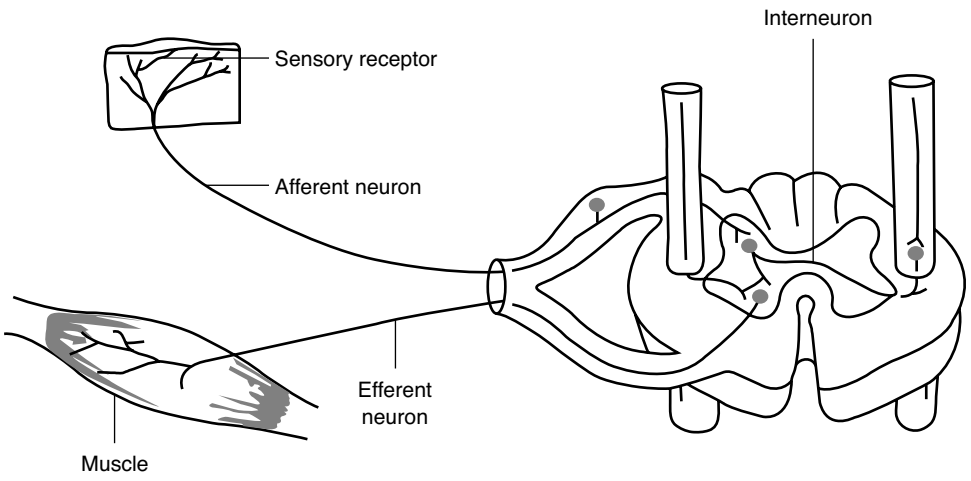


Figure 2.6 The neural connections in a spinal reflex. A reflex is an example of innate procedural knowledge

food and avoid getting sick. The neural mechanisms behind such preferences are probably not too complex either. They need involve little more than a mapping between the odor and the emotional response.

Given the above examples, we see that it is not so far-fetched for us to be born with procedural knowledge. This knowledge is in the form of simple neural circuits that map stimulus inputs to appropriate behavioral outputs. This knowledge can even be represented using the conditional rules we talked about in Chapter 1. A coding of a smell preference might look something like: "If *smell is fish* then *disgust*." The odor, if it satisfies the first part of the conditional, would then trigger the response in the second part.

But how did these circuits get there in the first place? The early nativists and rationalists either did not specify the source of innate knowledge or attributed it to God. Evolutionary psychology offers us another explanation. It attributes such capacities to generations of selection pressures acting on a species. These pressures promote the development of adaptive (survival-related) cognitive abilities. Evolutionary psychologists further argue that these innate abilities are **domain-specific**, meaning they are attuned to perform special operations only on a certain type of information. Domain-specific mechanisms can be distinguished from content-neutral or domain-general learning mechanisms, such as the processes of associationism proposed by Locke. Evolutionary psychologists can be considered modern-day nativists. See the evolutionary approach chapter for more on their views.

The phrasing of the nature-nurture debate as allowing for only a single alternative, either one or the other, is misleading. Although some traits may indeed be the product entirely of nature or of nurture, there is a large middle ground consisting of traits or cognitive abilities that can result from the complex interaction of the two. In these cases, nature may set constraints or limits on environmental influence. Take memory for example. Tsien et al. (1996) engineered a mutation in a gene that affects a particular type of receptor in the hippocampus, a brain area responsible for the learning of new information. Rats with the mutation did poorly in a memory task as compared to normal rats in a control group. Tang et al. (1999) did something even more remarkable. Through genetic manipulation they increased production of a particular subunit in the hippocampal receptor. This change increased the receptor's effectiveness. Rats with this "enhanced" version of the gene outperformed rats with normal receptors on a spatial memory test.

This research is exciting because it shows that memory in these animals is at least partially under genetic control. However, it is also well documented that human memory capability can be improved through organization and the use of memory strategies (Roediger, 1980). The way in which these genetic and environmental factors interact to determine memory in any given individual is complex. It could be that any amount of organization could produce no more than a small increase in memory performance if the aforementioned gene were lacking. Alternatively, those with the enhanced version of the gene who also employ memory strategies could perhaps acquire "supermemories," and then no longer need to employ memory strategies in order to remember effectively.

The Mystery of Consciousness

Consciousness is a complex concept and has no single agreed-upon definition. In its broadest sense, we can think of it as the subjective quality of experience (Chalmers, 1996). It may be thought of as our individual subjective awareness of mental states. These states include sensation, perception, visual images, conscious thought processes, emotions, and sense of self, just to name a few. But these states assume a person is in a normal, awake, and alert frame of mind. The issue becomes more complex when we think of other types of consciousness; for example, being unconscious, asleep, in a drug-induced state, hypnotized, or meditating. There are clinical cases representing other states of consciousness as well. In dissociative identity disorder, a person can alternate between separate personalities. Each personality can possess unique skills and may or may not be aware of the others. In split-brain patients, one half of the

brain can possess an awareness of an object the other half does not possess. For simplicity, we do not consider these alternate states of mind.

An interesting aspect of consciousness is whether it is unitary or divided. Subjectively, our consciousness seems to be unitary. That is, one recognizes himself or herself to be one person, experiencing things in the present moment. When one studies the brain, though, one finds that there is no single place or even time where consciousness seems to happen. Instead, the brain in action is a case of activity going on all over the place. Furthermore, the brain may even be processing different aspects of a single experience at different times. How can we reconcile this objective evidence with our subjective experience? See the In Depth section for one theory on this apparent contradiction.

Chalmers (1996) makes a distinction between phenomenal and psychological concepts of mind. The **phenomenal concept of mind** is essentially the idea of mind as a conscious experience. Mental states in this view need to be explained in terms of how they feel. The **psychological concept of mind** sees mental states only in terms of how they cause and explain behavior. Here, mind is characterized by what it does—how it feels is irrelevant. Philosophers have concerned themselves primarily with the former, psychologists and cognitive scientists with the latter. To make this distinction clear, imagine biting into a candy bar. A phenomenal investigation would attempt to explain why you experience the mental states of "sweetness" or "chocolate" and why you might perceive them differently than somebody else. A psychological investigation would concern itself with the neural circuits that become activated during the bite, how they might be represented computationally, and how this explains when you might stop eating. In this section, we concern ourselves with the phenomenal concept of mind and its relation to consciousness, since the psychological view is in most cases the topic of the remainder of this book.

The What-It's-Like Argument

Nagel (1974) says there is "something that it is like" to have a conscious mental state. When you bite into a candy bar, you have a subjective conscious experience of tasting it. The candy bar of course has no such experience. There is nothing that "it is like" for the candy bar being bitten. This is one way of describing consciousness—that organisms that possess it can be described as having some sort of experience. Things incapable of supporting consciousness cannot.

But what is this experience like? Nagel asks us to imagine what it must be like for a bat to navigate by echolocation. In echolocation, the bat emits high-pitched sounds. The sound waves bounce off an object in the animal's path and

Figure 2.7 What is it like to be a bat?

the animal uses the reflection time as a measure of the object's distance (see Figure 2.7). We could conceivably build a machine that could compute echolocation the same way a bat does. It might even perform as successfully as the bat. But this would not tell us what it is like for the bat to experience the world in the way it does. We have seen this argument before in our evaluation of functionalism. There we said that a functional description of a cognitive process does not account for the qualia, or subjective experience, of the process.

The problem here is that science can only provide an objective account of a phenomenon and consciousness is an inherently subjective state. As organisms capable of supporting consciousness, we can introspect and analyze what it is like to have or to experience a mental state. Unfortunately for cognitive science, this is not what is needed. Cognitive science must instead have a scientific and objective account of what consciousness is. Frank Jackson (1982) aptly illustrates the difference between objective and subjective accounts of a conscious experience. He asks us to think about a neuroscientist named Mary who is well trained in the physical mechanisms underlying color vision. She understands everything there is to know about how the eye and brain process color information. Mary, however, is colorblind. Imagine now that we take away her

colorblindness and allow her to look at a tomato. Interesting questions arise. Does Mary learn anything new by this experience? Does the scientific community gain anything by Mary's (or anybody else's) description of what it is like to see red? Jackson argues that we do gain something, and that science needs to explain this new information. This gulf between an objective and a subjective description of mental phenomena goes by the name of the **explanatory gap**.

In contrast to this position, some state that subjective knowledge is not factual knowledge at all and therefore does not constitute any kind of an explanation. Personally knowing what it is like to taste a candy bar or to see red is not the same thing as objectively and factually knowing it. Adopting this position, we as investigators would be forced to ignore introspection and any other form of subjective description. Our focus would be only on legitimate objective techniques for studying the mind, such as experimentation and brain scanning.

Mind as an Emergent Property

Consciousness is a "hot" topic in contemporary cognitive science. In the past fifteen or so years there have been renewed interdisciplinary efforts to understand it. A number of well-known authors have published books for academic and layperson audiences that outline their definitions and views on the subject. These authors' theories are too numerous to mention here. We instead describe one popular theory in this section and a more detailed description of another in the In Depth part of the chapter.

John Searle (1992) introduces a new twist on consciousness in his book *The Rediscovery of Mind*. He argues that consciousness is an **emergent property** of the brain. An emergent property of a system is realized through the interaction of the system's parts. He says if we have a given emergent system S, made up of elements a, b, c, and so on, then the features of S may not be the same as the features of a, b, c, and so on. This is because the features of S arise from the causal interactions of the parts. Water for example has the features or properties of "liquidity" and "transparency." The H_2O molecules that make it up do not share these properties. The causal interactions of these molecules, that is, their bumping into each other, give rise to these properties. In the same way, Searle says, consciousness is a property of the brain but not of its parts. If we take neurons to be the relative parts, then they have their own properties, such as being able to communicate via electrical signals with one another. These properties that are inherent in the way the neurons interact give rise to consciousness, but the properties of individual neurons need not be those of a conscious mind.

Searle is very careful to point out that he is not a reductionist. He does not believe consciousness is reducible to its parts. In fact, emergence is just the opposite of reductionism. In reductionism, explanation goes downward and a phenomenon is directly explainable in terms of what is happening at a smaller scale. In emergence, explanation goes upward. The smaller now gives rise to the larger. The large-scale phenomena are more than just what is happening in and around the parts and cannot be explained solely by an account of what the parts are doing. This idea is similar to the concept of a gestalt in perception. Gestalts are discussed in the next chapter (The Psychological Approach).

Searle seeks to avoid the monism-dualism dichotomy of the mind-body problem. He does this by talking about consciousness as a property rather than a substance. He likens consciousness to an emergent characteristic of what brains do in the same way that digestion is what stomachs do, or photosynthesis is what plants do. He sees consciousness as a natural process and a byproduct of the brain's nature. However, he does classify conscious mental states as separate from physical ones. He states that they constitute a unique and novel category of phenomena, with an independent reality and a distinct metaphysical status.

Evaluating the Emergent View of Mind

As appealing as this formulation is, it still leaves us with some vexing questions. The reformulation of consciousness as a property, and a nonphysical one at that, still begs the question: what is a property? If a property is not physical, then of what substance is it? Although attempting to avoid the mind-body debate, Searle seemingly ends up as a type of dualist. Restating consciousness as a nonmaterial property of a material brain doesn't get us any further toward understanding what this type of property is. Also, it is not clear how emergence happens, that is, we do not yet have an understanding of the relationship between microscopic and macroscopic properties. In the case of water, we can say its properties have something to do with the three-dimensional shape of the H_20 molecules and other conditions, such as the surrounding temperature. For consciousness and the brain, this relationship between the microscopic and the macroscopic is far more ambiguous.

Searle's reason for believing in a nonmaterial consciousness is based on his conception of the difference between physical and mental things. For physical things, we can make a distinction between appearance and reality. A piece of wood may subjectively appear a certain way to us; as brown, as having a certain length and weight, and so on. These characteristics can also be measured

objectively; we can put the wood on a scale, use a ruler to determine its length and a wavelength detector to measure its color. For mental things this distinction between the subjective and objective goes away. As regards mental experience, Searle believes that appearance is the same as reality and that our subjective introspections are objectively correct. But if this were true, we would have to trust our intuitions about the mental world as metaphysically special and nonmaterial.

In opposition, Paul Churchland (1995) points out that this reliance on the infallibility of introspection is an outdated notion. He notes that introspection often does not give us direct and accurate knowledge of the mental. Our inner assessments of mental states can be quite often and notoriously wrong. It is commonplace for us to err in judging our thoughts, feelings, and desires. Many of the early psychologists relied on introspection as a means to study the mind. The next chapter provides a more elaborate discussion of the problems they encountered.

Consciousness and Neuroscience

What does the brain have to do with consciousness? Is there some part of the brain or some particular pattern of neural activity that gives rise to consciousness? What is the neural correlate of conscious experience? Although philosophers have been debating the relation between the brain and mental phenomena for millennia, recent advances in neuroscience have yielded more specific insights into these questions. Let's examine some of them here.

In general, the neuroscience view is that consciousness results from the coordinated activity of a population of neurons in the brain. Popper and Eccles (1981) see consciousness as an emergent property of a large number of interacting neurons. A different idea is that there are neurons specifically devoted to producing consciousness. Crick and Koch (1995) believe these are located throughout the cortex and in other areas associated with the cortex. Activity in at least some subset of these neurons produces conscious experience. They believe that these neurons are special and that they differ from other neurons in terms of their structure and function. A similar but slightly different conception is that any cortical neuron may contribute to a conscious experience; however, different groups of cortical neurons mediate different types of conscious experience.

If there were special consciousness neurons, where might they be located? It has been proposed that one area is the intralaminar nuclei of the thalamus (Purpura, 1997). The thalamus is a relay center for incoming sensory information.

It sends information from each of the different sensory modalities, such as vision, audition, touch, and taste, to specialized areas of the cortex devoted to processing the information. Lesions of or damage to this brain region result in coma and loss of consciousness. It may be that these thalamic neurons, because they have widespread projections to many cortical areas, serve to activate or arouse other cortical neurons. Specific activity in different cortical regions may then account for specific forms of consciousness. For example, activation or arousal in the occipital visual regions may correspond to visual awareness, while activation of the somatosensory cortex may produce awareness of different parts of the body.

Churchland (1995) formulates a neurocomputational theory of consciousness that focuses on connections between the intralaminar nuclei of the thalamus and disparate cortical areas. The circuit consists of ascending projections from the thalamus to the cortex, as well as descending pathways from the cortex to the thalamus. Figure 2.8 shows the anatomical layout of this area. These pathways are recurrent, meaning that a signal can be sent back and forth inside it. In this case information coming into the thalamus can be passed to the cortex, while the cortex can also pass information back to the thalamus. Recurrence is an important network property because it allows for feedback and learning. Recurrent activity in a network may sustain information over time and be the basis for conscious mental awareness. Recurrent artificial neural networks and their properties are described in Chapter 7 (The Network Approach).

Churchland believes the characteristics of this network can account for a number of different features of consciousness. One such feature is the capacity of consciousness to hold information over time—the equivalent of a short-term memory in which we are aware of the current moment in relation to the past. This is in keeping with the network's recurrent nature, since information can be held over time as it is cycled back and forth. Churchland also shows that this network can maintain activity in the absence of sensory inputs, for example, when we are daydreaming or thinking with our eyes shut. It can additionally explain why we lose consciousness during sleep, why it reappears during dreaming, and a host of other such features.

Churchland is quick to acknowledge, however, that it is the dynamical properties of this recurrent network and not its particular neural locus that make consciousness possible. He admits that a consciousness circuit may exist in places that have been suggested by other researchers. One such area is within the right parietal lobe (Damasio, 1994). The neuroscientist Rodolfo Llinas has suggested that consciousness may arise within the layers of the primary sensory cortex itself. He has written a comprehensive book that lays out his views (Llinas, 2002).

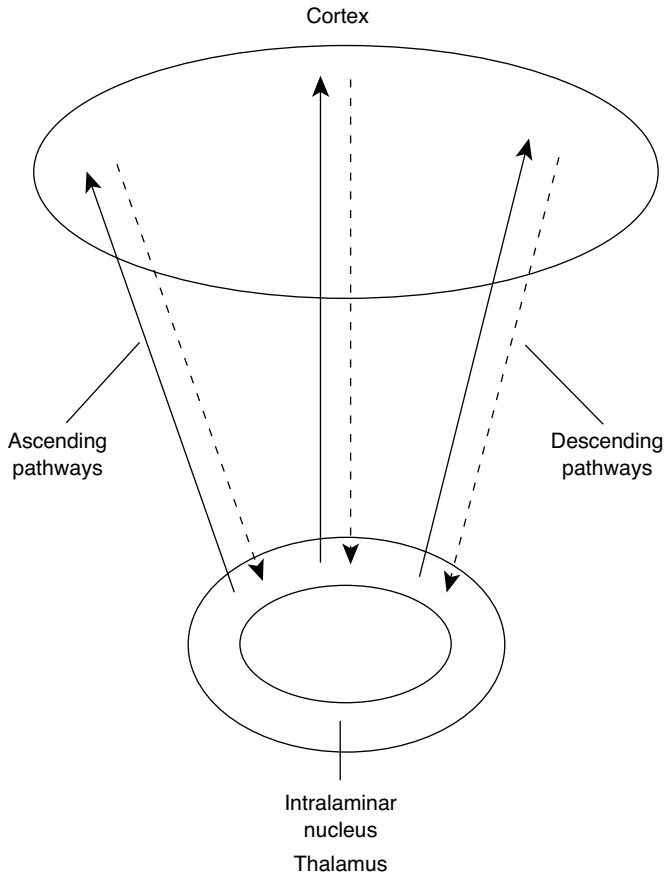


Figure 2.8 The consciousness circuit proposed by Paul Churchland. Activity in these pathways may give rise to conscious experience

Consciousness and Artificial Intelligence

Researchers in artificial intelligence (AI) design algorithms to perform real world computational tasks such as language comprehension or problem solving. Many of these algorithms can be judged successful from a behavioral standpoint because they adequately perform their tasks, some under a variety of different conditions. If we define thought as computation in some physical substrate, as functionalists do, then we can also without much risk say that

these programs are "thinking." But are these programs conscious, or could they ever become conscious? This is a much riskier proposition, since, as we have seen, consciousness implies more than computation. It seems to involve subjective experience and perhaps other things. In this section, we address the question of whether a machine can be conscious. This is perhaps the most interesting philosophical issue (but not the only one) in artificial intelligence today.

There are a variety of different perspectives on whether or not a machine can become conscious (Freedman, 1994). These may be generally classified into two categories. The **strong AI** view asserts that consciousness can arise from a purely physical process. Followers of this perspective believe that, eventually, as we create machines with greater complexity and computational power, we will see consciousness emerge in them. Proponents of **weak AI** claim that consciousness is itself either not a physical process and so can never be reproduced, or is a physical process but such a complex one that we will never be able to duplicate it artificially. We elaborate on what is meant by strong and weak AI later in the chapter on artificial intelligence.

Let us examine the arguments both for and against strong AI. Daniel Dennett (1998) raises several points in its defense. He mentions that many phenomena that used to have mystical and supernatural explanations now have scientific ones. Consciousness should be no different, he argues. Some have claimed that consciousness may only be possible in an organic brain. Dennett concedes this may be true, but notes that science has already been able to mechanically reproduce small-scale biochemical processes. An alternate counterargument is that consciousness is simply too complex to be artificially replicated. In response to this Dennett says that consciousness of a more basic form may not require a sophisticated artificial substrate. Dennett ends by noting any conscious machine will probably have to develop this capacity through an extended learning process, just as humans do. From a practical standpoint this is not a barrier, since a number of machines that learn from experience have been designed.

Perhaps the most persuasive and well-known argument against the strong AI position is the **Chinese room scenario** (Searle, 1980). In this hypothetical situation a man is in a room by himself. Outside the room is a person who asks a question in Chinese. This question is converted into written Chinese symbols on paper. The man in the room understands no Chinese whatsoever, but has a book of rules that tells him how to relate the Chinese symbols that make up the question into a set of symbols constituting a response (see Figure 2.9). These written symbols are then converted back into a spoken reply. For example, if the outside person utters "How are you?" in Chinese, the man

Figure 2.9 Searle's Chinese room scenario

in the room may, using the rule book, counter with "I'm doing fine!" To an outside observer, it would seem as if the person in the room understands Chinese. After all, he has a reply to any question that is given. But Searle's point is that the man knows no Chinese. He is only following a prescribed set of rules that maps one set of symbols onto another. This is a rote execution of an algorithm and, according to Searle, it is all a machine can do. Therefore, he says, machines can never "understand," "know," or "be aware of" the information they process. They cannot be conscious of what they do. Consciousness of the human sort requires something more than just following an algorithm.

To Searle, these extra ingredients are intentionality and meaning—aspects of mental representation discussed in the introductory chapter.

Boden (1990) raises a number of objections to the Chinese room argument. First, the terms understanding and intentionality are not well defined. Understanding could be operationally defined as being able to respond successfully when asked, rather than as entailing some inherent meaning on the part of the person. Second, a person who remained for some time in the Chinese room or even a machine in the room, if of sufficient complexity and if left there long enough, might eventually achieve some level of understanding. Either might eventually notice that certain combinations of characters always go together and from this learn the basic elements of syntax. Finally, Boden mentions that one could take this argument to its extreme by building a robot that, following rules, is indistinguishable from a human, yet one would have intentionality and the other would not.

Overall Evaluation of the Philosophical Approach

One of the main advantages of the philosophical approach is that it allows us to ask much broader questions than those of other disciplines. A cognitive psychologist studying memory for nouns might wonder why concrete nouns are recalled better than abstract ones. This psychologist is constrained into formulating specific questions and hypotheses by the narrow focus of the research. This very focus of course is an advantage, since it allows the researcher to examine and understand a natural phenomenon in depth. A philosopher examining the results of this same research is free to inquire about the nature of concreteness or what it means that something is abstract. He or she could also inquire as to how concrete or abstract stimuli are processed in other cognitive systems, such as attention and language. Of course he or she is free to ask even more fundamental questions, like: Why do we have memory? What purpose does memory serve? What would a person be like without a memory? Philosophy thus shows us the "bigger picture." It gives us key insights into the relationships between different areas of study, within and between disciplines, and therefore plays a very important role in the interdisciplinary endeavor of cognitive science.

Keep in mind that philosophy is a non-empirical approach. It does not utilize the scientific method. Concepts in philosophy are validated through logical reasoning and argument rather than by systematic observation and experimentation. For this reason, the conclusions reached in philosophy are speculative and

theoretical until tested. Philosophy is better suited to the asking of important questions—how, what, and why we should study something—rather than to providing definitive answers. These answers come from the scientific disciplines. It is therefore important that a good two-way working relationship exist between philosophers and these science-based approaches.

In Depth: Dennett's Multiple Drafts Theory of Consciousness

Dennett, in his book *Consciousness Explained* (1991), outlines an interesting theory on the nature of consciousness. He begins by refuting the classical view of consciousness. The classical view, promoted by Descartes, posits a single point in the brain where all information funnels in. This area is a supposed center of consciousness, where we experience the world or the contents of our thoughts in a coherent, unified way. Dennett calls this center the "Cartesian theater." It is as if our consciousness is the result of a projector displaying information onto a movie screen. The individual sitting in the theater watching the screen then has a single conscious experience of what is playing. Figure 2.10 gives a representation of the Cartesian theater.

There are a number of problems with the Cartesian theater. To start, linked modes of information do not arrive within the brain simultaneously. Light from an event precedes the arrival of sound. The sight of a fireworks display reaches the mind prior to the sound of the explosion, yet we experience the two in unison. This suggests that our consciousness is constructed; the visual experience is kept in check or delayed until arrival of the sound, at which point the two are integrated into a unified percept of the fireworks. This example and others imply that consciousness does not occur in real time—but (in many instances) several fractions of a second or so after an event. Our experience of consciousness as direct and immediate seems to be an illusion.

Another problem with the Cartesian theater is that, anatomically, it is difficult to find a brain region that links incoming sensory input and outgoing motor outputs. There is no central processing unit (CPU) in the brain as there is in a computer. The task of a computer's CPU is to schedule and coordinate ongoing activity. Furthermore, the Cartesian theater analogy requires an observer in the audience watching the screen. This observer is the subjective self who experiences the screen's contents. But how is this person inside our heads interpreting the image and having the conscious experience? To explain this, we need to posit another mechanism or theater inside this person's head with another even smaller person and so on, ad infinitum. This is known as the

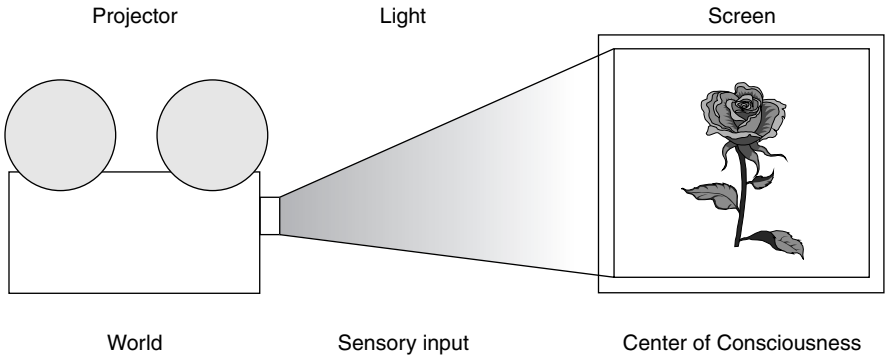


Figure 2.10 The Cartesian theater explanation of consciousness

homunculus problem in psychology and philosophy. **Homunculus** translated means "little man." An effective theory of consciousness must avoid the logical conundrum of homunculi nested inside each other.

Dennett replaces this problematic formulation with a multiple drafts model of consciousness (see Figure 2.11). In this model mental activity occurs in parallel. Rather than projecting to a single location for processing in unison, different ongoing streams of information are processed at different times. Each of these streams can correspond to different sensory inputs or thoughts. Processing or editing of the streams can occur, which may change their content. Editing can consist of subtractions, additions, and changes to the information. Awareness of a stream's content can happen before or after editing takes place. To illustrate, take our fireworks example. One mental stream would contain the visual experience of the fireworks while another would contain its auditory representation. The visual stream would undergo editing in the form of a delay to synchronize it with the auditory stream. Then the information from both streams could be tapped to produce awareness.

There is abundant evidence in support of the multiple drafts model. Take, for instance, the organization of the visual system. It adopts a "divide and conquer" strategy. The visual system carves up different aspects of an object during pattern recognition. These aspects are each processed separately in different parts of the brain by anatomically distinct pathways. This information is later combined to yield a unitary percept, but we are not aware that the information has been separated and then united.

A famous experiment by Loftus and Palmer (1974) also provides support for Dennett's idea. In this study, participants viewed a film of a car crash.

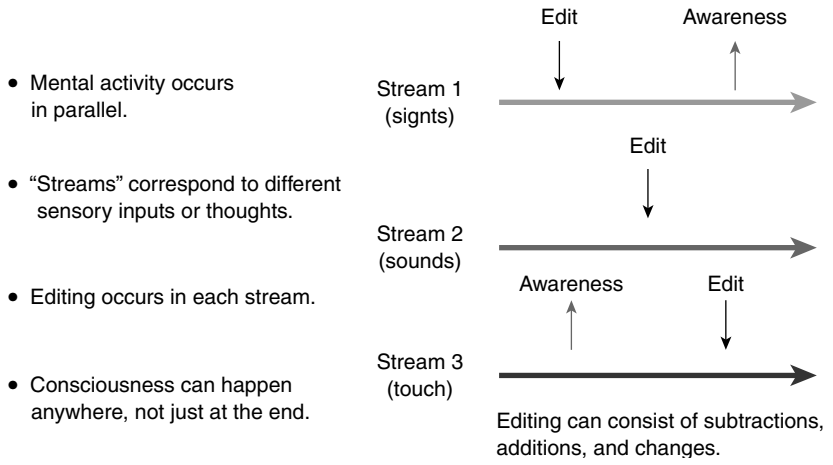

Edit
Awareness
• Mental activity occurs in parallel.
Stream 1 (signts)
Edit
• "Streams" correspond to different sensory inputs or thoughts.
Stream 2 (sounds)
Awareness
Edit
• Editing occurs in each stream.
Stream 3 (touch)
• Consciousness can happen anywhere, not just at the end.
Editing can consist of subtractions, additions, and changes.

Figure 2.11 Dennett's multiple drafts model of consciousness

Afterward, they were asked to estimate the speeds of the cars. The crucial manipulation was in how the question was asked. Some were asked how fast the cars were going when they "bumped" into each other. Others were asked how fast they were going when they "smashed" into each other. As you might predict, those who were queried with a mild descriptor like "bumped" estimated that the cars were moving more slowly. Those queried with a more severe descriptor like "smashed" estimated the speeds as considerably higher. These results suggest that our memories of an event are not perfectly preserved "snapshots" of what happened, but are actively edited over time. The posing of a question and other subsequent experiences after the event can cause the memory to be edited and changed.

Dennett's theory also allows for different levels of awareness. Some information that is part of a stream may be available to conscious awareness and could be verbally described by the individual experiencing it. Other data streams we may only be vaguely aware of, but they can persist and influence additional mental processes. Yet other information may simply fade into the background. We may never be aware of this information. These three levels of awareness are comparable to Freud's conscious, preconscious, and subconscious aspects of mind, discussed in the next chapter.

In summary, Dennett's theory is more logically coherent and captures some of the empirical evidence on conscious experience. It suggests there is no central place where consciousness happens, but that multiple mental events occur

in parallel. These events may be edited and changed in such a way that consciousness need not take place in real time. We may or may not be aware of these events.

Minds On Exercise: Decision Making

Join up with a small group of fellow students. Everybody in the group should agree on a list of ten possible graduate or other schools to which they might apply. Along with the list, provide information on such things as prestige/reputation, geographic location, student-teacher ratio, library resources, and so on. Next, each person in the group will design an algorithm that will enable him or her to pick a winning school. This algorithm should be a formal representation of the mental process he or she would go through in making that decision. Once everyone is finished, they can show and explain their algorithms to the rest of the group. Be sure to discuss the following questions:

1. Compare and contrast the different decision making processes. Are any similar? Can they be put into different categories? On what basis?
2. Did two people use different algorithms but pick the same school? Did two people use the same algorithm but pick different schools?
3. Would the same algorithm pick the same schools each time? If not, in what ways might an algorithm change over time? What does this say about predictability and the free will–determinism debate?

Food for Thought: Discussion Questions

1. Are you a monist or dualist? Take one side of this debate and argue in support of it. Can cognitive scientists legitimately adopt a dualist position?
2. The billiard ball model of causality implies that with a sufficient understanding of a system we should eventually be able to predict any natural phenomenon. Is this true? Will we eventually be able to perfectly predict the weather? Human behavior?
3. What kinds of knowledge are we born with? Create a list of human knowledge that you think is innate. Is all of this knowledge procedural? Can we be born with factual, declarative knowledge?

4. Describe some emergent properties that exist in the natural world. Can these tell us anything about how the mind might emerge from the brain?
5. Searle argues that consciousness is much more than a machine executing an algorithm. Is this sufficient for consciousness, or is something more required? What might this something more be?

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Bechtel, W. (1988). *Philosophy of mind: An overview for cognitive science*. Hillsdale, NJ: Lawrence Erlbaum.

Chalmers, D. (1996). *The conscious mind*. Oxford, UK: Oxford University Press.

Churchland, P. M. (1986). *Neurophilosophy: Toward a unified science of the mind-brain*. Cambridge, MA: MIT Press.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul: A philosophical journey into the brain*. Cambridge MA: MIT Press.

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science*. New York: Oxford University Press.

Damasio, A. (1994). *Descartes' error: Emotion, reason and the human brain*. New York: Putnam.

Dennett, D. C. (1991). *Consciousness explained*. Boston, MA: Little Brown & Company.

Dennett, D. C. (2003). *Freedom evolves*. New York: Viking.

Hofstader, D. R., and Dennett, D. C. (1981) *The Mind's I*. Basic Books.

Llinas, R. R. (2002). *I of the vortex: From neurons to self*. Cambridge, MA: MIT Press.

Pinker, S. (2002). *The blank slate: The modern denial of human nature*. New York: Viking.

Rand, A. (1963). *For the new intellectual*. New York: Signet.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

3

The Psychological Approach: A Profusion of Theories

"Mind, n. A mysterious form of matter secreted by the brain. Its chief activity consists in the endeavor to ascertain its own nature, the futility of the attempt being due to the fact that it has nothing but itself to know itself with."

—Ambrose Bierce, 1911

What is Psychology?

This chapter highlights a number of different perspectives adopted within the discipline of psychology, primarily during the 20th century. Each has a relatively unique answer to the question: What is mind? But first we need to ask ourselves another question: What is **psychology**? As a discipline it is best defined as the scientific study of mind and behavior. Psychology uses the scientific method as a means of gaining valid knowledge. Its focus of study includes internal mental events such as perception, reasoning, language, and visual imagery. It also however studies behaviors, which are external events. Behaviors include such things as walking, talking, and running. This distinction between mind and behavior makes its appearance in the fundamental questions this approach attempts to answer. These questions include: What are

the contents of the mind? How do these contents interact with each other? How does the mind explain what we do?

The different movements in psychology described in this chapter have focused on different themes. Many of the movements, including voluntarism and structuralism, were concerned with cataloging the "stuff" inside our heads. That is, they tried to list the basic elements of mind and to delineate how they interact to form other elements. We see this question of mental content taken up later by other approaches. Cognitive psychologists who studied memory later on tried to describe concepts that exist in our memories and how these concepts are related to one another. They formulated the idea of a mental lexicon or dictionary as a way of explaining how concepts exist in memory. In the network approach, we see that semantic networks were created to explicitly describe the location and arrangement of concepts.

A second major theme that emerges in the history of psychology is centered on operations—what the mind does, rather than what it contains. We see the issue of operations being addressed first by the functionalists. Functionalism veered away from a listing of the mind's supposed "parts" and studied the ways in which the mind performed various mental actions. The psychoanalytic psychologists like Freud were also operational in their approach. Their focus was on how mental structures such as the id and the ego dynamically interrelate. Gestalt psychology focused on a specific aspect of mental operations, namely, how mental parts form wholes. The Gestalt psychologists wanted to know how the mind creates larger-order structures during perception and problem-solving. This theme of mental operations occurs repeatedly in other areas of cognitive science. Cognitive psychologists would later form elaborate models of mental processes to explain perception, memory, and attention. In neuroscience, brain scans would reveal the neural sequences of events that underlie mental tasks. Connectionists would devise neural networks with specific patterns of operation to simulate mental function, and in the domain of artificial intelligence, computers would be programmed with complex sets of rules to mimic the kinds of things the brain can do.

Interestingly, in the behaviorist movement we see a counter-reaction against the whole concept of studying the mind. Behaviorists viewed the mind as something that passively mapped aspects of the environment onto an organism's response. It was the environment and not the mind, they believed, that controls a person's actions. The study of behavior and its relation to the environment appears again in cognitive science in the field of robotics. A challenging goal in this field is to get a machine to perform some task successfully in a real-world environment. To do so requires an understanding of how variable stimulus inputs can map onto a robot's possible responses.

A number of commonalities run through the different movements in psychology. Many were influenced by developments in other branches of the sciences. Particularly influential were chemistry and physics. Voluntarism and structuralism both adopted fundamental ideas from chemistry and embraced the idea that the mind, like the physical world, was made of basic elements that combine into larger wholes. Similarly, the Gestalt psychologists adopted the idea of field theory from physics and employed the notions of fields, particles, and forces in their descriptions of mental phenomena. In addition, many movements in psychology arose as counter-reactions to existing movements or as a means to address problems raised by prior movements. Both functionalism and Gestalt psychology, for instance, arose in opposition to structuralism.

Psychology and the Scientific Method

Before describing the foremost theories of each movement we need to digress and examine how psychology goes about its business. Early psychologists relied on techniques such as introspection and phenomenology. Modern psychologists employ a wider variety of methods. They administer questionnaires and surveys, analyze case studies of single individuals, and record behavior in the wild through naturalistic observation. A number of modern day psychologists also cross over into other disciplines such as neuroscience or artificial intelligence and so employ the methodologies of these approaches as well. Many of these techniques are characterized by the principles of science.

The scientific endeavor in general is characterized by the hypothetic-deductive approach. In this approach a hypothetical conjecture about the way the world works is tested deductively. The testing is accomplished by carefully observing the way the world works under controlled conditions. If the observation supports our conjecture, we can elaborate and expand upon it. If it doesn't we must change it to account for what has been observed. The testing, as we alluded to above, can assume a variety of forms. In neuroscience, it can be the observation of brain-damaged patients or the scanning of brain activity. In artificial intelligence or cognitive psychology it may involve constructing a computer simulation. The scientific method, in which an experiment is conducted to test a hypothesis, is perhaps the most widely used method in all of psychology and cognitive science. For this reason we will look into it now.

The **scientific method** uses an experiment, which is designed to test a **hypothesis**, a specific statement about the world. For example, one could hypothesize that a participant's memory for a list of words will be worse when the participant is listening to music. The validity of a hypothesis is based on

the outcome of an experiment. The results of an experiment can either support or fail to support a given hypothesis. In this example, if it were the case that memory was worse while the participant listened to music, we could say the results do support the hypothesis. Hypothesis testing in turn helps researchers construct a **theory**, a more general understanding of the world that organizes a set of facts and aids us in understanding how the world works. The data of this particular experiment might support a theory that distraction interferes with our ability to remember information.

Any experiment involves the use of independent and dependent variables. An experimenter manipulates an **independent variable** to see if it will produce a change. Going along with our previous example, a researcher might vary the presence of music in a memory experiment. In the music condition, participants memorize a list of thirty words while listening to music being played in the background. In the no-music condition, a different group of participants must memorize the same thirty words without any music playing. The **dependent variable** is what is measured or observed by the experimenter to see if a change of some kind has occurred. Such a change may be thought of as the effect or thing that happened as a result of the independent variable. In this instance, the independent variable would be music and the dependent variable might be the average number of words successfully recalled.

Generally experiments consist of a minimum of two conditions. The **experimental group** receives the independent variable while the **control group** does not. In our hypothetical example, the music condition would correspond to the experimental group and the no-music condition to the control group. Assuming that everything else that makes up the two conditions is held constant by the researchers, then any difference between the two sets of results must be attributed to the manipulation. If we found that average word recall was higher in the no-music condition than in the music condition, then we could conclude that background music interferes with memorizing words. As stated above, this would support our hypothesis. It should be noted that historically psychology has employed other techniques. Among these are introspection and phenomenology, which will be defined later.

Mental Atoms, Mental Molecules, and a Periodic Table of the Mind: The Voluntarist Movement

Duane Schultz and Sydney Ellen Schultz, in *A History of Modern Psychology,* 4th ed. (1987), give a good overview and critique of the various theoretical movements in the history of psychology. These movements include voluntarism,

Figure 3.1 Wilhelm Wundt established the first experimental psychology laboratory in Leipzig, Germany, in 1879

structuralism, functionalism, behaviorism, Gestalt psychology, and the psychodynamic view. They all preceded the cognitive revolution that we will discuss in the next chapter. We refer our readers to their text for a more detailed discussion of the history and ideas behind these movements.

We will begin our discussion of psychology with voluntarism. The **voluntarism** movement viewed the mind as consisting of elements, but stressed that these elements were assembled into higher-level cognitive components through the power of the will. It was the will, or voluntary effort of the mind, that was seen as the force behind the creation of more complex mental elements. German physiologist and psychologist Wilhelm Wundt (1832–1920) was the founder of voluntarism (see Figure 3.1). Another field in science, chemistry, had a significant influence on voluntarism. During Wundt's time chemists were attempting to describe the material world in terms of its basic components and

how they combine. The Russian chemist Dimitri Mendeleev developed the periodic table of chemical elements during this time. The idea behind this table was that the entire physical universe consisted of atoms, characterized by different properties, and that atoms could, under certain conditions, combine to create more complex, higher-order molecules. Wundt may have also been attempting to create a periodic table of mental elements, and to specify how these elements combine (Marx & Hillix, 1979).

Other historical antecedents influenced Wundt. His focus on elements is something he inherited from 19th century British empiricists and associationists, although he differs from them in that in his view it is the will, rather the laws of association, that synthesizes and creates novel elements. His approach is also analytic and reductionistic in that he believed the contents of consciousness could be reduced to basic forms. Wundt's methodology was shaped from the experimental methods in use in the sciences. In fact, voluntarism heralds the beginning of psychology as a formal discipline, since it is the first time in history that mental phenomena were systematically studied in accordance with the scientific method. The official birth of psychology is usually given as 1879, the year Wundt opened his laboratory in Leipzig, Germany.

Wundt's method was **introspection** or internal perception. *Introspection* literally means "inward looking." Just as one can look out at the external world to see various objects, such as a chair or table, Wundt believed one could also look inward to experience and describe mental objects. He presented the students in his lab with various stimuli, such as colored shapes, and asked them to introspect. The students then recorded their subjective experiences in relation to the stimuli. Although various forms of introspection had been used by philosophers for centuries, Wundt attempted to systematize and objectify the technique. He had his students put themselves in a ready state of attention prior to their introspecting and repeat their observations several times as he varied specific physical aspects of the stimulus, such as size and duration of the exposure. This kind of methodological exactness exemplifies the influence of the scientific method.

Wundt believed psychology should study consciousness. However, he distinguished between two types of conscious experience. **Immediate experience** is our direct awareness of something. For example, if we see a rose, our perception of the rose as red is immediate. It is the redness we experience directly while we are looking at it. If someone then asked us what we were looking at and we responded "a red rose," then that thought would be a mediate experience (it was a thought about the rose). **Mediate experiences** are those that come from mental reflection about an object. Wundt emphasized the study of immediate experiences. He believed they are the best way to describe the basic elements of mind, since they are "untainted" by elaborate thought processes.

Wundt went on to develop a **tridimensional theory of feeling**, according to which all feelings can be characterized by three dimensions. These dimensions were pleasure-displeasure, tension-relaxation, and excitement-depression. Wundt would play a metronome, a device that produces audible clicks at different time intervals. As he varied the rate of the metronome, he discovered that some rhythms were more pleasant than others. He also reported a feeling of tension that accompanied his waiting for a click, followed by a feeling of satisfaction after the click occurred. Finally, he noted feeling excited when the tempo of clicks was increased and calm when they were played at a slower rate. Based on these introspective observations, he believed any feeling could be characterized by pleasure-displeasure, tension-relaxation, and excitement-depression.

Although voluntarism attempted to describe and classify the basic elements of mind, it also needed to account for the fact that we perceive unitary wholes, rather than collections of individual elements. For instance, when looking at a face, we see a face in its entirety, not just a collection of its different parts, two eyes, a nose, and a mouth. To account for this, Wundt postulated the principle of **creative synthesis**, also called the law of psychic resultants. According to this principle, the mind actively organizes disparate elements together such that the resulting whole contains new properties. These new properties cannot be explained by the characteristics of the individual elements themselves. A similar phenomenon is seen in chemistry. Here, characteristics of individual atoms, for example, hydrogen and oxygen, are insufficient to specify the characteristics of the water molecules that form when they combine. Water has unique properties that cannot be understood simply by analyzing it down into its component parts.

To Wundt, creative synthesis was an active process in which the mind took hold of elements and forcefully combined them. This contrasts with the earlier views of the empiricists and associationists, who saw mental combination as a passive and mechanical effect. They argued that mental wholes were created through automatic processes of association, and that these processes did not require the active participation of the mind deemed necessary by Wundt. This topic of the relationship between parts and wholes does not end here. The Gestalt psychologists would have much to say on this issue several years later.

Evaluating the Voluntarist Approach

The voluntarist approach was beneficial because it was the first scientific attempt at studying the mind. The establishment of a laboratory, the application of experimental methods, and the formulation of the clearly defined goals of listing elements and describing their combinations are all strengths

of this movement. However, many of the criticisms that have been leveled at voluntarism also center around the elements of its methodology, introspection in particular. Critics have pointed to a number of flaws in introspection. First, a mental experience may change over time. A person's experience of red may undergo alteration after several seconds have elapsed—the experience perhaps becoming less vivid. Also, it may be impossible to separate immediate and mediate experiences because the act of introspecting could itself change an experience. According to this idea, the simple reporting of red, even as a sensation, involves reflective and perhaps other thought processes that alter the true experience of red. And there was the issue of individual differences. Some observers in Wundt's lab experienced the same stimulus in different ways, which suggested that people's expectations or learning experiences could change their perceptions. Finally, Wundt was never able to compile a short list of fundamental mental elements comparable to the list of elements that had been compiled in physical chemistry. The list of elements derived from his students' introspections with respect to their immediate, sensory type experiences was growing far longer than what a simple periodic table would accommodate.

Structuralism: What the Mind Is

Structuralism shares a number of ideas with its predecessor, voluntarism. The subject matter of psychology was again conscious experience, and the method was a modified version of introspection with an emphasis on the scientific method. There were still major theoretical differences between the two schools. The structuralist view of the mind held that, once again, the mind was a passive agent, with mental elements combining according to mechanistic laws. Structuralism, per its name, focuses on mental elements; that is, the structure of mind is to be understood in terms of basic elements and their combination—suggesting once again the analogy to chemistry. Its progenitor was U.S. psychologist Edward Bradfort Titchener (1867–1927).

Titchener cautioned against making the **stimulus error**, which is to say, confusing our true experience of an object with a description of the object based on language and past experience. When one looks at a banana, the true experience would be that it is yellow and curved—not the recognizing of it as a banana or a type of fruit. This distinction parallels Wundt's differentiation between immediate and mediate experience. Titchener in addition broke away from Wundtian-style introspection. He believed that only well-trained observers could introspect accurately and not make the stimulus error.

According to Titchener, psychology had three goals. These were: (1) to describe consciousness in terms of its simplest and most basic components; (2) to discover the laws by which these elements or components associate; and (3) to understand the relation between the elements and their physiological conditions. The initial aim was, as with voluntarism, to come up with a set of fundamental mental units. Titchener believed a mental element was fundamental when it remained constant over numerous introspective trials. So, if multiple introspective observers all consistently experienced an object in the same way, then this experience qualified as a true element.

The combination of elements, in Titchener's scheme, was not effected through active mental processes, as Wundt believed. Instead, Titchener saw the mind as a passive mechanism or substrate within which elements combined according to set laws. In chemistry a **reagent** is a substance added to a mixture in order to produce a particular chemical reaction. Reagents are used to quantify these reactions and their chemical products. In a similar fashion, structuralists believed the mind to be a reagent, a medium inside of which mental reactions and processes unfolded. Researchers of the day even went so far as to call their subjects reagents!

Like Wundt, Titchener listed his elements. He described a total of 44,000 sensation elements alone. Out of these, 32,820 were visual and 11,600 were auditory (Titchener, 1896). He thought that each of these elements was fundamental and indivisible, but that each was capable of combining with others to form more complicated perceptions and ideas. Titchener believed all sensations could be characterized by four attributes. **Quality** is what separates a sensation from any other sensation. The experience of heat therefore has a different quality than that of sound. **Intensity** refers to how strong a sensation is; for example, a noise can be loud or soft, a light bright or dim. **Duration** refers to how long a sensation persists, whether it is short-lived or longer lasting. Sensations are characterized also by **clearness**. Sensations that one pays attention to possess greater clarity. To these basic four, Titchener later added **extensity**, the extent to which a sensation fills or occupies space. The sensation of pressure coming from a pencil tip that touches the skin has less extent than that that would come from a chair-bottom.

Evaluating the Structuralist Approach

Many of the advantages and disadvantages of the structuralist movement are identical to those of its ancestor, voluntarism. Structuralism further refined scientific methodological procedures and applied them to the study of psychological

phenomena. It employed a variety of techniques, including measurement, observation, and experimentation. Introspection in a modified form would be put to use later on in other areas of psychology, such as psychophysics, in which observers judge the relative intensity of stimuli, and clinical psychology, in which participants are asked to evaluate themselves along personality continua, such as those centering on anxiety or depression (Schultz & Schultz, 1987).

More elaborate critiques of introspectionism were now pointed at the structuralist school. Among these were that the process itself is subjective and unreliable, that training participants to introspect only biases their responding even more, and that some mental experiences, such as habits, happen without conscious awareness and so cannot be introspected about at all. The psychoanalytic school would expound considerably on this idea of an unconscious mind that is inaccessible to attention or awareness. In conclusion, structuralism was found to be too analytic and reductionistic. It was found to overemphasize the role of low-level mental elements and to ignore holistic perception and experience, once again opening the way for the Gestalt psychologist's concern with this issue.

Functionalism: What the Mind Does

As is so often the case in scientific disciplines, a theoretical perspective that has been entrenched for many years is often quickly replaced with an alternative, sometimes completely contrary, perspective. This was certainly the case with functionalism's supplanting of structuralism. Rather than focusing on the contents of mind, **functionalism** instead focused on what the mind could do. Its emphasis was the mental processes or functions that operate on the elements, instead of the elements themselves. Harvey A. Carr (1873–1954), one of the later U.S. functionalists, summarizes the subject matter of functionalist psychology as follows:

> Psychology is primarily concerned with the study of mental activity. This term is the generic name for such activities as perception, memory, imagination, reasoning, feeling, judgment, and will. . . . Stated in comprehensive terms, we may say that mental activity is concerned with the acquisition, fixation, retention, organization, and evaluation of experiences and their subsequent utilization in the guidance of conduct. (Carr, 1925, p. 1).

The formal development of psychological functionalism in the United States is credited to John Dewey (1859–1952) and James Rowland Angell (1869–1049),

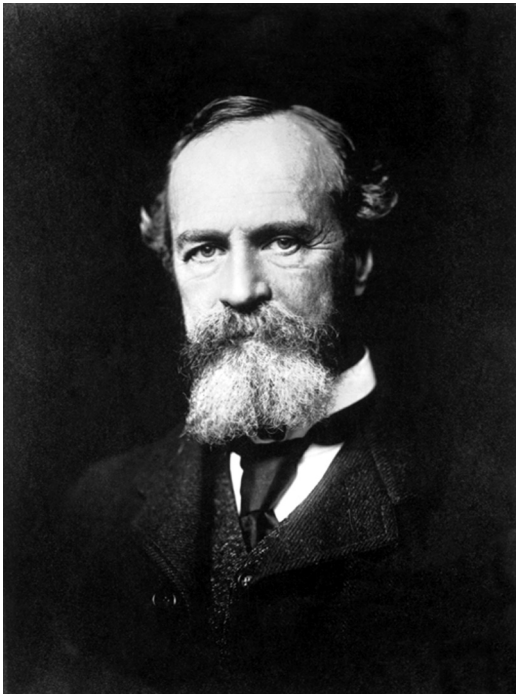


Figure 3.2 William James was one of the early American psychologists

Source: © CORBIS

as well as to Harvey A. Carr. However, it was William James (1842–1910) who was its pioneer and perhaps most lucid expositor (see Figure 3.2). James is often considered the greatest American psychologist. James rebelled against the Wundtian and structuralist conception of experience as being made up of discrete elements, and believed such elements existed simply as a result of the attentive, introspective process. He thought an individual erroneously created the notion of an element by attempting to "freeze" or stop a moment in the mind's ongoing activity. He stated that one person's subjective element in response to a perception, for instance, does not guarantee that the same element will exist in anyone else's mind who experiences the same perception—what he called the **psychologist's fallacy**.

James replaced the "mind as elements" notion with the idea of mind as a **stream of consciousness**, suggesting the mind was a process undergoing continuous flow or change. He likened thought to the water in a river that is always moving. He provided another analogy when, referring to mind, he declared: "Like a bird's life, it seems to be made of an alternation of flights and

perchings" (James, 1890, p. 243). James termed the resting-places or "perchings" of thought as *substantive*. **Substantive thought** occurs when the mind slows down, perhaps when focusing attention. James called the "flights" of mind **transitive**. They correspond to a less focused, more associative form of thinking.

James Rowland Angel articulated the three major themes of functionalism (Angel, 1907). The first theme was the study of **mental operations**. By this, he meant functionalism should investigate how a mental process operates, what it accomplishes, and under what conditions it occurs. Second, Angel believed functionalism should study the **fundamental utilities of consciousness**. One ought to understand the role consciousness plays in the survival of the organism. This means asking such questions as: Why are we conscious? Of what utility or usefulness is consciousness? How does consciousness help keep an organism alive? Third, functionalism should study **psychophysical relations**, the relations between the psychological mind and the physical body. The term also refers to the total relationship of the organism to the environment, and to how the organism, both physically and mentally, may have adapted to the environment.

From these themes it is clear that functionalism was strongly influenced by Darwin's theory of natural selection. Functionalism then sought to explain not just mental operations in and of themselves, but how the mind in general arose through evolutionary pressures to serve the organism. The idea of an organism-environment fit and how it may help to explain a variety of different psychological characteristics was to be elaborated on more fully at a later date by evolutionary psychologists. Functionalism thus serves as the theoretical precursor to evolutionary psychology, discussed in greater detail later in this book.

Evaluating the Functionalist Approach

The functionalist movement broadened the field of what could acceptably be studied in psychology. Unconscious phenomena were fair game, as was the study of children, the mentally retarded, and the "insane." It allowed for a wider variety of methods, such as tests, questionnaires, and objective behavioral descriptions, and it permitted the continued use of introspectionism (Schultz & Schultz, 1987).

Criticisms of functionalism came, perhaps not surprisingly, from the structuralist camp. In 1913, C. A. Ruckmick , a student of Titchener's, accused the functionalist movement of failing to define the term "function," having noted that functionalists used the term in two different ways. First, function was used to refer to an activity or process itself, such as perception or memory. Second,

function was used to mean the utility or usefulness of an activity to the organism (Schultz & Schultz, 1987). Functionalists sometimes used these terms interchangeably, which sometimes invited confusion. Another reported fault of functionalism was that it was too applied and practical. Its adherents were accused of focusing excessively on the usefulness of functions. In contrast, the structuralists took a more basic scientific approach in their attempt to describe and elucidate the basic aspects of mind. This debate over which is better—basic or applied research—is ongoing in psychology and other sciences.

The Whole Is Greater Than the Sum of Its Parts: Mental Physics and the Gestalt Movement

In the early part of the 20th century, another field of psychology arose. This field was called Gestalt psychology and its three founders were all German. They were Max Wertheimer (1880–1943), Kurt Koffka (1886–1941), and Wolfgang Kohler (1887–1967). The Gestalt movement, like functionalism, was a counter-reaction against structuralism and the atomism it entailed. The Gestalt psychologists even referred to structuralism as "brick and mortar psychology." They saw wholes as more than just the sum of their parts; such an integrated whole they called a **Gestalt**. What characterizes the Gestalt approach is the importance of conscious wholes. The Gestalt psychologists believed that conscious wholes could not simply be reduced to a listing and description of their parts. The Gestalt psychologists also borrowed a metaphor from physics: they believed that mental parts combined into wholes in much the same way physical particles organized when subjected to fields of force.

The contributions of Gestalt psychology were greatest in two areas, perception and learning. Gestalt psychology is phenomenological in method. **Phenomenology** refers to subjective experience rather than objective description. When studying perception, for example, the Gestaltists preferred to create stimulus patterns, show them to observers, and have the observers describe their subjective experiences. Phenomenological description differs from introspection in that it focuses on a person's immediate and subjective perception of an external stimulus. It does not require training or an intensive examination of one's internal state. The Gestalt psychologists were looser methodologically when studying learning as well. They preferred to observe human or animal subjects find solutions to problems rather than set up rigorously controlled experimental conditions.

In vision, the relationship between parts and wholes is an important one. As mentioned earlier, when we look at an object, we tend to see the entire object,

rather than the parts of which it is composed. We tend to see a face, not an aggregate of eyes, nose, and mouth. We tend to see trees and not collections of leaves and branches. Max Wertheimer illustrates the problem this way: "I stand at the window and see a house, trees, sky. Theoretically I might say there were 327 brightnesses and nuances of color. Do I have 327? No. I have sky, house, and trees" (1923, p. 71).

Wertheimer then went on to formulate the **principles of perceptual organization**. These are ways in which visual parts group to form objects. The principles demonstrate that the relationships between parts, which can be independent of the parts themselves, are important to determining how these parts are assembled into wholes. According to the principle of **proximity**, parts that are close to one another in the visual field are perceived as a whole. Here, the physical distance between elements is a relationship, independent of those elements, but one that serves to group them together. According to the principle of **similarity**, parts that are similar in lightness, color, shape, or some other characteristic group together. The relationship here consists of the elements' shared properties. The principle of **closure** holds that parts that form a complete or almost complete and enclosed object go together. Wertheimer also proposed the idea of **pragnanz**, or "good figure." According to pragnanz, parts that are simple will group together. Simple organizations are those that have fewer parts or are bilaterally symmetrical. Wertheimer created **lattices**, or matrices of regularly arranged parts, to demonstrate these principles. Several lattices and other Gestalt stimulus patterns are shown in Figure 3.3. It should be noted that these are only a few representative examples of grouping principles proposed by the Gestaltists. More recent perceptual investigators have suggested others.

Wolfgang Kohler's greatest contribution was in the area of animal learning. Kohler studied chimpanzees on the island of Tenerife during World War I and described those studies in *The Mentality of Apes* (1927). He would place a banana on a ceiling hook and observe how the chimps would use available materials to try and reach it. The chimps at first would use different approaches at random—trying to knock the banana down with a stick, for example. He then noticed that the chimps, with great suddenness, often after a period of inactivity, would solve the problem by stacking crates on top of each other and climbing that stack of crates to reach the prize (see Figure 3.4). Kohler termed this behavior **insight learning**, the ostensibly spontaneous understanding of relationships that produces a solution to a problem. He believed that this phenomenon was operational in humans as well. For more on how humans exhibit insight learning, see the In Depth section at the end of the chapter. Because the solution is a holistic configuration and involves a set of interrelationships

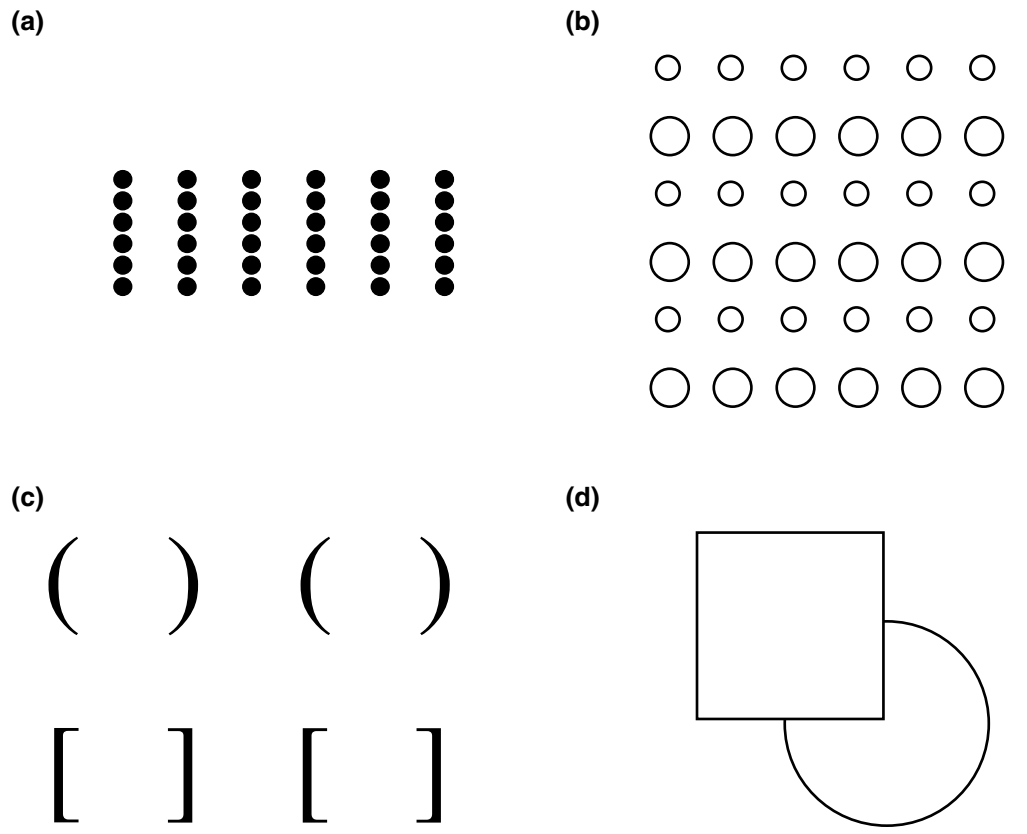


Figure 3.3 Dot lattices and other figures that demonstrate several of the Gestalt principles of perceptual organization. In the lattice in (a), the organization is seen in terms of columns, not rows, because the dots in that orientation group by proximity. In (b), the organization is one of rows, because the dots group by similarity of size. The inward pointing pairs of parentheses and bracket-like figures in (c) form a coherent whole because of closure. The shapes in (d) are almost always perceived as a square overlapping a circle, rather than a Pac-Man and a square, because a circle is simpler according to the law of pragnanz

among component parts, it demonstrates again the Gestaltist's emphasis on mental wholes. These wholes can be perceptual, as is the case during the perception of an object, or conceptual, as is the case during learning. It should be stressed that these wholes in the Gestalt view consist not just of individual parts, but additionally of the relationships between them.

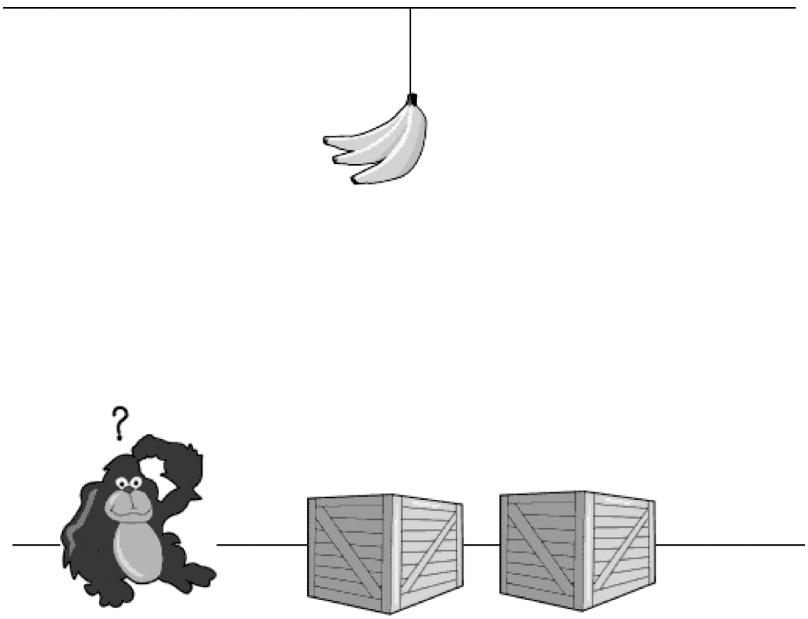


Figure 3.4 If you were the chimpanzee, how would you reach the bananas?

The Gestalt psychologists formulated a theory of how underlying brain mechanisms give rise to Gestalts. Wertheimer suggested the cerebral cortex was a dynamic system, in which the elements in force at a given time interacted. Kohler, in *Static and Stationary Physical Gestalts* (1920), further proposed that cortical processes operate like the fields of force that are studied in physics. Metal filings caught in an electromagnetic field generated by the poles of a magnet will organize themselves around the poles. These and related phenomena in physics are described by **field theory**. Similarly, it was believed that perceptual elements spontaneously organized themselves because they were caught in "mental force fields." In this mental equivalent of field theory, sensory impulses coming from a stimulus initiate neuronal activity. This electrochemical activity then produces a mental field that causes the perceptual grouping. Here again we see the influence on psychology of another discipline in the natural sciences.

The Gestalt psychologists believed there was a correspondence between the psychological or conscious experience on the one hand and the underlying brain experience on the other. This perspective is called **isomorphism**, because the perception is identical (Greek: *iso*) in form or shape (Greek: *morph*) to the brain activity that gives rise to it. The mental representation of elements and

the fields operating on them were thought to be quite similar to the conscious perception of the stimulus, in that same way that a map is similar to the geography it represents.

Evaluating the Gestalt Approach

Gestalt theory gave an interesting alternative formulation of the part-whole problem. It addressed the problem in a more detailed fashion than had the structuralists or voluntarists. It also spawned research traditions in perception and in learning and problem solving. However, critics soon pointed out a number of flaws. The Gestalt psychologist's methodology was the first aspect of Gestalt theory to receive harsh treatment. The phenomenological approach was labeled as "soft" and lacking in scientific rigor. Gestalt psychologists were accused also of being too theoretical and of failing to back up their assertions with empirically derived data. Much of their data, because it was not obtained within controlled experimental settings, was not amenable to statistical analysis. Their approach was thus critiqued as too qualitative and not quantitative enough in orientation. Field theory was additionally accused of being speculative and based on poorly defined physiological assumptions.

Some of the specific findings of the Gestalt researchers were also attacked. The principles of perceptual organization were said to be merely descriptive; they described how parts grouped, but did not provide a real explanation. Field theory as an explication for grouping was seen as too vague as well as inadequate. Critics averred that the organizational laws seemed sufficient to describe grouping when one principle was at work, but failed to make predictions about organization taking place under more complex real-world circumstances. Figure 3.5 shows three dots, A, B, and C. Dots A and B group by proximity while dots B and C group by similarity. What is the resulting organization? A further criticism was that the concept of pragnanz was ill-defined. Stimuli can be constructed where a more complicated, not simpler organization is

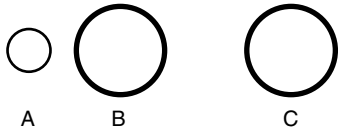


Figure 3.5 The first two dots group by proximity. The second and third dots group by size similarity. The resulting perceptual organization cannot be predicted by the individual laws of organization alone

perceived. Gestalt findings in learning were also not immune from such evaluations. Kohler's idea of insight learning as something that happened quite rapidly was questioned. Later investigators found that insight learning in animals such as chimpanzees does not always occur suddenly, and can depend on factors such as prior experience and learning (Windholz & Lamal, 1985).

Mini-Minds: Mechanism and Psychoanalytic Psychology

Psychoanalytic psychology, sometimes called psychodynamic psychology, has come to occupy a less prominent position in the contemporary academic world. During its heyday it was quite influential, particularly as a means of understanding and treating psychological disorders. The founder of the movement is perhaps one of the most well known names in the discipline: Sigmund Freud (1856–1939). Freud's basic framework was modified subsequently by a number of others, later called **neo-Freudians**. These included Carl Jung (1875–1961), Alfred Adler (1870–1937), and Karen Horney (1885–1952).

Psychoanalytic psychology as promulgated by Freud saw the mind as made up of distinct components, or "miniature minds." Each of these minds competes with the others and vies for control of behavior. Psychoanalytic psychology posited not just one state of consciousness, but three, and emphasized the role of the unconscious mind, of which the individual has little awareness and over which he or she has little control, in influencing thought and action. Freud also thought that sex, pleasure, aggression, and other primitive motivations and emotions were powerful influences on personality, as were early childhood experiences.

Freud proposed a three-tiered system of consciousness. The three tiers were: (1) the **conscious** mind, which contains those thoughts and feelings of which we are aware and can directly access; (2) the **preconscious** mind, that is, those aspects of mind that we can bring into awareness with effort; and (3) the **unconscious** mind, or the aspects of mind of which we are completely unaware. An example of conscious content would be knowing the street address where one lives. Preconscious content requires some mental effort for one to be made aware of it. An example would be trying to recall and eventually recalling what one did on his or her birthday last year. Unconscious content may never come into awareness, although one goal of psychoanalytic therapy is to try to allow this to happen. An example of unconscious content are memories of childhood abuse or other traumatic experiences from the early developmental years.

Freud described three other mental structures, each having a different operating principle. The **id** contains unconscious impulses and desires such as sex

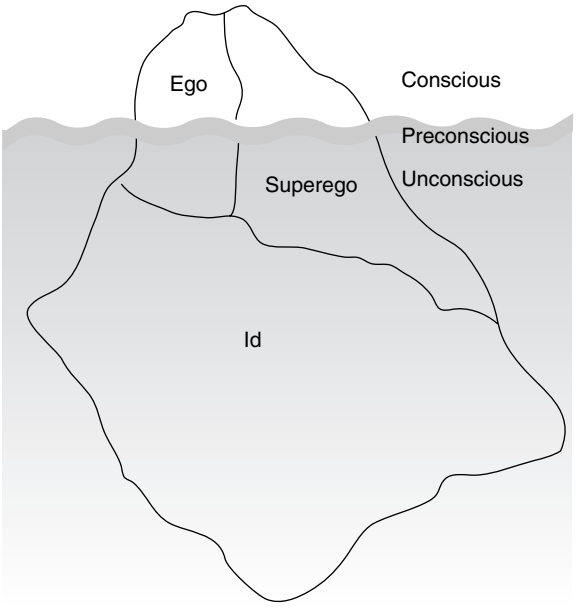


Figure 3.6 Sigmund Freud's iceberg model of mind

and hunger. It operates on the **pleasure principle** and attempts to attain gratification for its desires immediately. The **superego** is responsible for our ethical sense. It operates on the **idealistic principle** and motivates the individual to do what it considers morally sound or proper. The **ego** balances the competing demands of the id and superego. It operates on the **reality principle**, spurring one to act in a rational and pragmatic fashion.

A metaphor Freud proposed for thinking about these structures was an iceberg, depicted in Figure 3.6. He likened the conscious mind to the visible part of an iceberg, above the waterline; thus it is always visible. The preconscious mind is equivalent to the part of the iceberg near the waterline, since it is sometimes submerged and unconscious, but sometimes above the waterline and inside awareness. The largest part of the iceberg is then the unconscious, which is always submerged and hidden from view. The id lies completely in the unconscious, its large mass indicating its power. Both the ego and the superego occupy all three states of consciousness.

A crucial part of this perspective is the dynamic interplay between the id, ego, and superego. The id demands immediate satisfaction of its appetite for pleasure, while the superego in most cases advocates controlling or suppressing

such urges in the name of decency. It then falls to the ego to attempt a rational and viable solution to the conflict. The ego is thus being barraged with the id's impulses and the superego's restraints. If it fails to satisfy either one, the result is anxiety. To shield itself against anxiety, Freud believed the ego constructs **defense mechanisms**, which reduce or redirect anxiety in various ways. Examples of defense mechanisms include **repression**, the banishing of anxiety-arousing thoughts and feelings from consciousness, and **sublimation**, the transformation of unacceptable impulses into socially valued motivations. If a person lost memory of his or her being abused as a child, Freud would say that he or she repressed the memory because it was too painful. If instead that person channeled the anger he or she felt as a result of having been abused toward becoming successful in a career, he or she would be sublimating.

Mechanism and determinism are inherent in the psychodynamic view. One can envision Freud's model of mind as an elaborate machine with different interacting parts. In his writings, there are many terms drawn from mechanics, electrical circuitry, and hydraulics. Without too much imagination, it is possible to construct a Freudian hydraulic or "plumbing" model of the mind, with pumping stations connected by valves and tubes. Id impulses in this model could be represented as an increase in water pressure. Forces deriving from other components that release or redirect this pressure could counteract that increase. This view of mind as machine implies determinism as well. Freud thought all mental events, including slips of the tongue and dreams, were determined and that nothing in human thought or behavior could occur by chance.

Evaluating the Psychoanalytic Approach

Many of the ideas in the psychoanalytic approach have permeated the wider culture. Freud drew our attention to important issues such as the unconscious and the influence of biological forces. The psychoanalytic approach served to stimulate further research in these areas and inspired several generations of clinical practice, but its lasting legacy may be in the strong counterarguments it elicited.

The neo-Freudians, although they accepted most of the substantive parts of Freud's theory, differed in several regards. They placed more emphasis on the role of the conscious mind and downplayed the importance of sex and aggression as all-powerful motivators. Modern psychological research suggests Freud overestimated parental and early childhood influence and underestimated the importance of other social factors, such as peer influence (Frieze et al., 1978). In addition, repression has been found to be a relatively rare mental response to psychological trauma. In fact, it is typically the case that horrible memories

are remembered quite clearly (Loftus, 1995). The unconscious mind is now viewed not so much as a seething caldron of primitive urges, but as a reservoir of information processing that goes on without awareness (Greenwald, 1992).

Psychoanalytic psychology has also been critiqued for its scientific shortcomings. The theory is not based on objective observations. Its raw material comes from Freud's subjective notes about his patients, written several hours after therapy. Freud additionally offers few hypotheses that can be scientifically verified or rejected. Some have censured psychoanalytic descriptions as being incapable of being proved one way or another. For example, the presence of anxiety in an individual could indicate fear of giving into id impulses. The absence of such anxiety could indicate its repression. The perspective thus offers after-the-fact explanations for personality. It also fails to predict behavior or traits (Myers, 2001). One could in this framework explain why someone is angry, but not anticipate accurately when he or she might get angry in the future.

Mind as a Black Box: The Behaviorist Approach

The weaknesses of the psychoanalytic position were the behaviorist's strengths. Instead of attempting to describe complex mental operations, the behaviorists, as their name implies, focused entirely on the study of behavior. In their view, the internal workings of the mind were simply too complex as well as incapable of being measured in an objective and scientific fashion. Behaviors, however, which are the actions that are produced by organisms, such as running, grasping, or bar pressing, are external, which makes them easily measurable and quantifiable. They are thus ideally suited for scientific study. Behaviorism was influenced by work in animal and comparative psychology. In these disciplines, there was a trend to discover which environmental conditions might cause animals to act in certain ways, as opposed to an emphasis on constructing elaborate mental models of how they might be thinking. Humans were de facto lumped into the same category as animals, making behaviorism a more general science, and an extension of the more rigorous natural sciences.

The behaviorists clearly saw themselves as true scientists. Because the mind could not be studied scientifically during their time, they preferred not to study it at all! This did not mean that behaviorists such as Skinner denied the existence of mind or that they thought brains were unworthy of study. Rather, they believed the scientific method could not be appropriately applied in this situation and so redirected their empirical investigations to something that was, namely, behaviors.

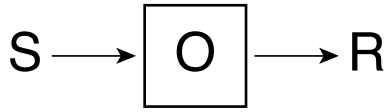


Figure 3.7 The behaviorist stimulus-response model of behavior. The mind of an organism (O) that emits a response (R) to a stimulus (S) in this view could not be well understood. The mind was treated as a "black box."

Behaviorism did reject outright the study of consciousness through introspection, its adherents claiming that the goal of psychology was not to study consciousness at all, but instead behaviors and the antecedent conditions that give rise to them. Figure 3.7 shows a version of the stimulus-response (S-R) model that represents their position. In the model, a stimulus (S) in the environment impinges on an organism (O). The stimulus then causes the organism to produce a response (R). The most important class of stimuli that influence responses, according to behaviorists, are rewards and punishments. In this case, a stimulus might consist of a reward given to an organism, for example, a rat, which would then cause the rat to respond by continuing to press a bar. Notice that in this model, the mind, contained in the organism, is not depicted. It is treated as a "black box," meaning that it is unfathomable and unexplainable. The mind becomes simply an entity that serves to convert stimuli to responses.

John B. Watson (1878–1958) was the first person to lay out the assumptions and foundations of the behaviorist movement. He distinguished between implicit and explicit responses. An **implicit response** is one that occurs inside the animal. It includes such events as visceral movements, glandular secretions, and nerve impulses. An **explicit response** on the other hand is overt and directly observable and includes actions such as talking, reaching, and so on. Although Watson allowed for the study of implicit responses in the realm of mental activity, it was difficult to investigate them, he observed, due to the limited instrumentation available at the time. The focus of behaviorism as a result centered on the study of explicit responses.

Other important figures associated with the movement were Edward Lee Thorndike (1874–1949), Ivan Pavlov (1849–1936), and Burrhus Frederick Skinner (1904–1990). Pavlov was a Russian physiologist who studied the dog's digestive system. In his lab, he noticed that dogs would salivate not only when eating, but also immediately before. This was true only for dogs that had been fed in the lab for some time. He assumed that, in addition to the food itself, any stimulus that was put forward repeatedly just before the food was presented

would also be able to trigger salivation. To test this idea, Pavlov applied meat powder to the dogs' tongues. This caused them to salivate. Then, he sounded a tuning fork each time, just prior to administering the powder. He did this over several trials. Finally, he sounded the fork without presenting any powder and found that the dogs still salivated.

As a result of these experiments Pavlov was inspired to propose several terms. The **unconditioned stimulus** (UCS) is that which elicits a response all on its own. In this instance, it is the meat powder. An **unconditioned response** (UCR) is a response elicited by an unconditioned stimulus, here the salivation caused by the administration of the meat powder. A **conditioned stimulus** (CS) elicits a response only after it has been paired repeatedly with an unconditioned stimulus, here the sound of the tuning fork. Last, a **conditioned response** (CR) is a response elicited by a conditioned stimulus, the salivation caused by the tuning fork. This form of learning came to be known as **classical conditioning**. Note that it pertains only to built-in reflexes and does not require any conscious thought on the part of the animal. The dog learns simply by associating the two stimuli.

B. F. Skinner was responsible for uncovering more about the nature of learning. Whereas classical conditioning works only with involuntary reflexive behaviors, such as the salivatory response, operant conditioning is pertinent to any voluntary motor act and is thus far more general. **Operant conditioning** is that type of learning in which a behavior is strengthened if it is followed by reinforcement, and diminished if followed by punishment. A **reinforcement** is anything that increases the frequency of a response. Food is a good reinforcer for most animals. A child, if given a cookie for taking out the garbage, will be more likely to take out the garbage in the future. **Punishment,** on the other hand, is any consequence of a behavior that is apt to decrease the future incidence of that behavior. A painful stimulus such as electric shock can serve as an effective punishment. Lab rats repeatedly shocked for touching an object in their cage will soon cease to do so. Skinner was able to use various forms of reinforcement and punishment to control the behavior of animals such as rats and pigeons.

Evaluating the Behaviorist Approach

The strength of the behaviorist position is its adoption of a completely objective science of behavior. Behaviorists understood that mentalistic concepts were poorly defined and could only be measured via subjective methods such as introspection. They were very rigorous in their application of the scientific method to the study of psychological phenomena. As such, they were confident that their results yielded the most accurate and valid information. The

testimony to behaviorism's success is that it was the dominant paradigm in research psychology for approximately 50 years. It didn't begin to decline until it was challenged by the rise of cognitive psychology in the 1960s. Some of the specific challenges posed by the cognitive movement will be discussed in the next chapter.

Behaviorism had allowed for the controlled and systematic study of behavior, but the behaviorists' abandonment of consciousness and mind was the movement's greatest weakness. During the behaviorists' reign, evidence came to light that animals do indeed have mental representations that affect how they learn. The work of Edward Chace Tolman (1886–1959), a behaviorist himself, challenged some of the assumptions of traditional behaviorist doctrine. Tolman allowed rats to explore a maze on their own and in some cases transported them through the maze. In both scenarios, the rats were subsequently able to navigate the maze successfully, even though they had never been rewarded for doing so. Tolman suggested the rats developed a **cognitive map**, or mental representation of the maze, that enabled them to navigate the maze without conditioning. This shows animals can acquire behaviors through experience, without being subject to any reinforcement—a concept called **latent learning**.

Tolman believed in five initiating causes of behavior. These were the environmental stimuli (S), physiological drive (P), heredity (H), previous training (T), and age (A). Behavior (B) can then be expressed as a function of these in a simple equation: $B = f_x$ (S, P, H, T, A). But between these initiating causes and the final behaviors exist what he called **intervening variables**. It is the intervening variables that are the determinants of behavior. They are the internal processes that connect the prior stimulus situation with the response. An example of an intervening variable is hunger. Although hunger cannot easily be observed, it can be related to other experimental variables that are readily observable, such as the amount of time that has elapsed since an animal last ate, or how much food the animal consumes at the end of a period of not eating. Intervening variables were the behaviorist's attempt to objectify internal mental states. However, these states were only acknowledged to exist as they pertained to some external, easily measured variable, and in many cases it was difficult to objectify them. Tolman himself later came to accept that a complete understanding of intervening variables was perhaps not possible.

Overall Evaluation of the Psychological Approach

We have covered a plethora of theoretical positions in this chapter. There are several reasons for this. Psychology was the first discipline to systematically

apply experimentation to the study of mind. It was thus a new discipline with many followers advocating many positions. Psychologists additionally had a very difficult task in front of them, which was to try to understand things that at the time could not be easily seen or measured. This lack of precision and early reliance on nonscientific methods such as introspection may have led to an overreliance on theory. Also, as was not the case in other disciplines, there was no overarching theory or framework for psychologists to work within. It would not be until the rise of the cognitive approach and the adoption of an information processing perspective that some kind of unity would come to the field. We next turn our attention to this cognitive approach.

In Depth: Insight Learning

According to Wallas (1926), insight learning happens in four stages. The first is **preparation** and consists of the acquisition of and understanding of the problem as well as preliminary attempts at solving it. The second stage is **incubation**, whereby the problem is put aside for a while. During this period, there are no conscious attempts at problem solving, but the unconscious mind may be attempting to find or has found a solution. In the third stage, **illumination** occurs. It is a flash of insight, a sort of "Aha!" experience in which the solution comes suddenly to awareness. Finally, there is **verification**, in which the insight is confirmed and one checks to see that it yields a correct solution.

At least one experimental study supports this theory (Silveira, 1971). In Silveira's experiment, participants were presented with a chain-link problem, illustrated in Figure 3.8. Go ahead and try to solve this problem on your own now, as we will provide the solution below. There were two overall conditions. In the control group, participants worked continuously for one half-hour on the problem. Only 55% obtained a correct solution within this period. Members of the experimental group were exposed to one of four conditions; the results are shown in Table 3.1. Participants were given either short or long preparation periods and a short (one half-hour) or long (four hour) interruption. The interruption here is like an incubation period, as participants were allowed to do something else, but not to think about the task. In this experiment, the lengths of the preparation and interruption periods were the two manipulated independent variables, while the percentage of subjects to solve the problem correctly was the dependent variable. The results showed that only 55% of the participants came up with the solution under the short preparation conditions, regardless of the length of interruption. Those subjects who were operating under the long preparation/short interruption condition

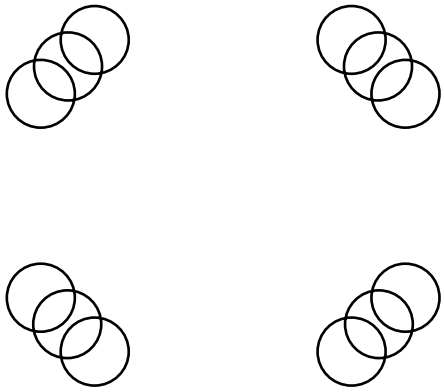


Figure 3.8 In the chain link problem, one starts with four pieces of chain with three links each. It costs $2 to open a link and $3 to close a link. Initially, all links are closed. The goal is to join all twelve links to form a circle and spend no more than $15

Table 3.1 The results of the Silveira (1971) experiment

	Preparation	
Interruption	*Short*	*Long*
Short (1/2 hour)	55%	64%
Long (4 hours)	55%	85%

improved slightly, with a success rate of 64%. It was the long preparation/long interruption condition that showed the best performance, with a success rate of 85%.

We can draw several conclusions from this study. The incubation, whether short or long, has no effect in instances in which people are allowed only a short time to familiarize themselves with the problem and to try to solve it. In these two cases, the 55% success rate is the same as that of the control group, who were given no incubation time at all. Problem solving improves slightly with a long preparation and a short interruption, but the greatest improvement is realized when both periods are long. Why is this? Incubation effects are thought to be the consequence of the unconscious mind's attempts at solving the problem. If an unconscious solution is obtained, it eventually enters conscious awareness.

Long preparation times could hypothetically give the unconscious greater familiarity with the problem. This, combined with the long incubation, also allows the unconscious more time to work on the problem, or, if a solution is obtained, has allowed it to percolate up into conscious awareness.

Gestalt psychologists were already aware of similar findings. In a much earlier study, subjects allowed to complete a series of tasks were found to have poorer recall with respect to those tasks than those who had been allowed to complete some of the tasks but not others (Zeigarnik, 1938). An example of one such task might be reading a story in its entirety, or reading the story but being interrupted part way through. This phenomenon of remembering more information when there has been an interruption has been dubbed the **Zeigarnik effect**. It was initially explained in terms of a tension system. Finishing a task was thought to produce a sense of completion and a reduction in tension. Failure to finish a task on the other hand produced an increase in tension, which perhaps yields the result of the information being thought about more, consciously or not, improving recall.

So here is the solution to the chain link problem. The standard incorrect solution involves opening and closing one link from each of the four pieces. This is incorrect, since it means opening four links and closing four links and spending $20. The correct solution requires the opening of all three links from one piece and using these to join the remaining three pieces, then closing them. This involves three opened and closed links, totaling $15.

Minds On Exercise: Introspection

This activity requires two people. Have a friend pick a relatively unfamiliar object, such as a bottle-opener or cat's toy. The introspector must close his or her eyes and is then handed the object. He or she is instructed to experience the object in as many ways as possible, by manipulating it, smelling it, and so on. As the friend experiences the object, he or she should look inward and examine the perceptions and thoughts that come to mind. Have participants do this for several minutes, then allow them to open their eyes and at this point to experience the object visually, while continuing to introspect. When finished, the introspector then writes down a list of his or her experiences and answers the following questions.

1. Was introspection difficult? Did you find it hard to "look inward"? Why or why not?
2. Were you able to identify the object with eyes closed? With eyes open? Did your inner experience change once recognition occurred?

3. How many of your inner experiences were immediate experiences, corresponding to direct sensory input, for example, colors or textures? How many were mediate experiences, corresponding to higher-order thoughts, such as the name of the object or its function?

4. Is it difficult to tell the difference between immediate and mediate experience? What about the subjective experience of "curve"? Is this a purely perceptual idea, or is it learned?

Food for Thought: Discussion Questions

1. The voluntarists and structuralists drew many metaphors about the mind from chemistry, such as the idea of the reagent. Can you think of additional concepts from chemistry that may serve as fruitful mind metaphors? Examples might include isotopes, ionic and covalent bonds, and diffusion across gradients.

2. The structuralists distinguish five aspects of a sensation: quality, intensity, duration, clearness, and extensity. Do you think there are more? Introspect about a sensation. Can you come up with other basic aspects, or do they reduce to these five?

3. Important to functionalism is explanation of the functions of mind, or mental operations. Produce a list of possible mental functions. Does the mind search, compare, select, or divide? Is there a finite number of mental operations, or do functions suffer the same fate as elements, in that there may be too many to effectively list and categorize?

4. The Gestalt psychologists suggested certain rules for ways in which perceptual parts may combine into wholes. Can you think of other rules of perceptual organization? Might parts that occupy the same spatial region go together? Do such laws of perceptual organization tell us anything about how conceptual parts combine into wholes? In other words, might ideas combine in the same way as perceptual parts?

5. Sigmund Freud proposes the id, ego, and superego as three "miniature minds" that compete with each other for control of behavior. Can you think of other possible component minds? Under what principles might they operate?

6. The behaviorists thought most learning was the result of conditioning that occurs either through association or a system of rewards and punishments. Can you think of examples where you learned something in a different way? When studying for a test or memorizing information, do you learn without conditioning?

Go to the website:
http://www.psych.purdue.edu/~coglab/VisLab/demos.html

Explore the various illusions and other visual phenomena you see there. Answer the following questions: (a) What is an illusory or subjective contour? What factors affect the perception of illusory contours? (b) Another Gestalt grouping principle is common fate. What is common fate? Describe the phenomena based on the "Objects by Common Fate" demonstration. (c) Describe the Rubin (1915) vase/face illusion. What is meant by the terms "figure" and "ground"?

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Freud, S. (1962). *The ego and the id*. New York: Norton.
James, W. (1890). *The principles of psychology*. New York: Dover.
Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt.
Kohler, W. (1927). *The mentality of apes*. Berlin: Royal Academy of Sciences; New York: Harcourt Brace.
Myers, D. G. (2001). *Psychology*. New York: Worth.
Schultz, D. P., & Schultz, S. E. (1987). *A history of modern psychology* (4th ed.). Orlando, FL: Harcourt Brace Jovanovich.
Skinner, B. F. (1971). *Beyond freedom and dignity*. New York: Alfred A Knopf.
Watson, J. B. (1970). *Behaviorism*. New York: Norton.

4

The Cognitive Approach I: History, Vision, and Attention

"I sometimes worry about my short attention span, but not for very long."

—Strange de Jim, 1974

Some History First: The Rise of Cognitive Psychology

Early pioneering work in what would come to be known as cognitive psychology began in the 1950s. Later, in 1967, Ulric Neisser published the first textbook on this subject. Cognitive psychology is the newest of the major disciplines in the overall field of psychology and is currently its most influential. The impact of the cognitive approach can be measured in terms of how it has affected other psychological disciplines. There are now cognitive theories in social, developmental, and clinical psychology.

To really understand the cognitive perspective, though, we need to backtrack to our prior discussion of behaviorism. You may recall from the previous chapter that for many decades behaviorism was the dominant movement

in psychology. Behaviorists avoided the study of mental phenomena because they believed that they were too difficult to define or measure. They stuck instead to external, observable behaviors, which were more amenable to scientific scrutiny.

Cognitive psychology can be seen as a backlash or counter-revolution against the behaviorist movement. This becomes evident when one examines the basic assumptions of the cognitive movement (Ashcraft, 2002). Whereas the behaviorists avoided studying the mental world and in some cases may have denied its existence, the cognitivists firmly acknowledged the existence of mental processes and focused their investigatory attentions on them. Whereas behaviorism saw the mind as a passive organ that operated according to simple rules of conditioning, cognitive psychology saw the mind as active, as selecting information from the environment, relating it to prior knowledge, and acting on the results of such processing.

Let us look more closely at some of the reasons for the so-called cognitive revolution. There were three main reasons for the rapid growth of this new perspective. The first was the failure of behaviorism to account for findings in areas such as language acquisition. The second was the invention of new measuring devices to examine mental activity. The third was the rise of the computer and the widespread use of the metaphor of mind-as-computer.

B. F. Skinner outlined the behaviorist take on language in 1957 in his book *Verbal Behavior*. He believed a child learned language through reward. If a baby said "mama" in front of its mother, the mother would get excited, smile, and talk back to the child—all forms of positive reinforcement. This would reward the child for having made that utterance, which he or she would then do more often. This process would be repeated, with the child uttering more complex words and sentences, being rewarded each time for correct pronunciation or syntax.

The linguist Noam Chomsky soon critiqued this theory of language acquisition (Chomsky, 1959). Chomsky argued that the behaviorists gave us no good account of why children suddenly use language. He asked: Why should a child utter a new sound or word? The child can't be reinforced until after he or she has already said it. Yet all children around the world spontaneously utter the basic phonetic elements of language. This suggests that there is an innate mechanism for generating language and that this mechanism may not, as the behaviorists would have us believe, be under environmental control.

Another problem is that children often combine the "pieces" of language in new ways to create new meanings. A child, after having learned the words "cat" and "sofa," might say "cat on sofa," even though he or she had never been reinforced for arranging the two words in exactly this way in a sentence.

The tremendous flexibility demonstrated by children in creating new meanings couldn't be accounted for by a system of reward.

A second reason for the rise of cognitive psychology was the development of new measuring tools. During the behaviorist era, there were no ways of "peering inside the head." That is, one could not directly measure mental processes. After the start of the cognitive movement, new technologies that provided a more accurate picture of mental processes as they were occurring emerged. The new devices included positron emission tomography (PET), computerized axial tomography (CAT), and magnetic resonance imaging (MRI). These techniques are described further in the neuroscience chapter.

But perhaps the thing that contributed most significantly to the decline of behaviorism was the increased use of computers. Prior to the 1960s computers were constructed using vacuum tubes. Because of the size and number of vacuum tubes needed to perform computations, computers were quite large. In some cases, entire rooms were required to house them. The transistor was invented in 1947, but was not widely applied to computer design until years later. The transistor performed the same computational function as a large number of vacuum tubes, but was much smaller. This miniaturization allowed for the construction of correspondingly smaller and cheaper computers. The widespread presence of computers spurred psychologists to begin thinking more about them. Psychologists realized that the mind, like a computer, could be viewed as a device that represented and transformed information. The mind-as-computer metaphor was born. Computers thus accelerated the adoption of the information processing view, not only in psychology, but also more generally, in other cognitive science fields.

The Cognitive Approach: Mind as an Information Processor

So now that we know how cognitive psychology came about, what exactly is it? **Cognitive psychology** is the study of knowledge representation and use in human beings. It is concerned with understanding how people represent, process, and store information. According to Ulric Neisser, "cognitive psychology" refers to all processes whereby the sensory input is transformed, reduced, elaborated, stored, recovered, and used" (Neisser, 1967). The many verbs used in the preceding sentence give us a sense of the many possible information processing activities of the human mind.

Cognitive psychology differs from other approaches in cognitive science in that its focus is on human information processing (as opposed to animal or machine modes of information processing). Like the practitioners of many

other disciplines in psychology, cognitive psychologists adopt the scientific method as their primary tool of investigation. Hypotheses are tested by analyzing data that has been obtained from controlled experiments. However, cognitive psychology also supplements experimentation with modeling and computer simulation. A specific information-processing model of a mental ability can be run on a computer. The results can then be compared against data from human experiments. This is often a synergistic and iterative procedure. Parameters of the computer model can be refined so as to provide a better fit between the computer model and the empirical data. Aspects of the simulation can also yield insights that cause researchers to go back and design new experiments.

A defining characteristic of the cognitive approach is the way it represents human information processing. These processes are often conceptualized using a **process model**. In a diagram of a process model, boxes are used to designate each stage or step in an information processing sequence. Arrows that point toward or away from the boxes represent the flow of information among the stages. Many of the figures in this chapter show a process model that depicts a particular theory of human computation. Feel free to examine a few of them now.

Process models, in the classical view of information processing, carry two assumptions. First, they are assumed to be sequential, meaning that information that lies within one stage is processed before it is output to the next. Information cannot be processed simultaneously in multiple stages. Second, excluding inputs and outputs, processing that occurs within one stage is independent of processing that occurs within other stages. These assumptions were later challenged by the connectionist view of information processing, which adopts a radically different architecture as the basis of cognition.

Process models are a very important part of the cognitive perspective. They are a powerful conceptual tool for understanding human information processes. In fact, the remainder of this chapter and the next chapter are devoted to describing the major processing models that underlie various domains in cognition. These domains include, but are not limited to, perception, attention, memory, imagery, and problem solving.

Modularity of Mind

An assumption of many cognitive theories is **modularity of mind**. According to this idea, the mind is made up of innate, functionally independent modules. The boxes in the process models that are described throughout this chapter and in the next chapter can in many cases be viewed as modules. The evolved psychological mechanisms that are discussed in the evolutionary chapter can

be viewed in this way as well. Because modularity of mind is such an important and pervasive assumption in the cognitive perspective, let's take some time now to examine it further.

Jerry Fodor is the outspoken proponent of the modular view. In his book *The Modularity of Mind* (1983), he gives a general description of the role these putative structures play. In his view, information arriving from the outside environment passes first through a set of sensory transducers. These convert the information to a code that is suitable for processing by the modules, which are domain-specific, that is, able to handle information of a specific type only. The modules then convert the results of their operations into a common code that can be interpreted by other non-modular domain-general processors.

In the same volume Fodor lists the many characteristics of a module. Modules are hardwired, meaning that they cannot be constructed from more basic units. They are genetically determined, domain-specific, fast, automatic, stimulus-driven, and not subject to control by a central authority. Modules are mandatory: they are triggered into operation by the presence of the appropriate information. Fodor also states that modules are informationally encapsulated. By this he means that other mental processes can only have access to a module's output; they cannot influence or access its inner workings.

Evaluating the Modular Approach

Case studies of brain-damaged patients support the modular view. There are many case studies of patients who have suffered damage to a particular brain area as a result of a stroke or accident. These patients then suffer very specific deficits. In an aphasic patient, damage to one part of the brain might hamper the patient's understanding or comprehending language, whereas damage to another region might cause difficulties in his or her speaking or producing language. See the chapter on linguistics for more on these disorders. Similarly, there are patients with developmental genetic disorders in whom one language ability has been preserved while another has not. Individuals with Williams Syndrome suffer grave deficits in visuo-spatial cognition but are relatively unaffected with respect to the processing of face-specific information (Bellugi, Wang & Jernigan, 1994).

Skeptics counter these claims. They point out that in many brain-damaged patients there is not a clear-cut dissociation of one function from another (Tyler, 1992). Also, genetic disorders of the sort described above rarely evidence highly specialized deficits—usually more general impairments are manifest (Bishop, 1997). Evidence from studies of early brain development additionally fails to support claims of modularity. Experiments that have utilized brain-scanning

techniques show that some language abilities are bilateralized, that is, located in both cortical hemispheres, early in development. Hemispheric specialization occurs later, which strongly suggests that these abilities cannot be genetically preprogrammed (Mills, Coffey-Corina, & Neville, 1993).

Current thinking is that the brain may contain a number of processing structures that are modular, but that these may be a result of normal developmental processes. However, the view opposite to that of modularity, that the brain is entirely domain-general, does not appear to be tenable either. The task for future researchers in this area is to define which developmental factors shape the formation of modules and which cognitive functions are more general-purpose and non-modular in nature. The role that development and genes play in the creation of modules is discussed further in the evolutionary approach chapter.

Theories of Vision and Pattern Recognition

Perception is the process by which we gather information from the outside world via the senses and interpret that information. Most work in perception has focused on vision and secondarily on audition. Far less research attention has been given to the remaining senses. This bias is species-centric, as vision is our most sophisticated sense and the one we rely on most. Many of the major perceptual theories are therefore based on and framed exclusively in terms of visual processing. In this section we will examine several broad theories of human vision, as we acknowledge that there is much theoretical work that remains to be done for the other senses.

One of the main functions of a visual system is the recognition of patterns. To demonstrate how important this is: imagine looking out at the world and not being able to recognize what you see. You would be able to navigate successfully, for example, you would be able to step out of the way of an oncoming car, but you would not recognize what it was that just went by. You would be able to see your best friend, but not know who he or she is. You would be able to see writing, but not read it. **Pattern recognition** is the ability to identify objects in the environment. We seem to do it effortlessly. Yet, as we will see, this process is actually quite complex and far from being entirely understood.

Template Matching Theory

Any description of a pattern recognition process must begin with the stimulus. The stimulus is the actual object in the external world that we are trying to recognize. Light striking the stimulus is reflected and projects an inverted image of the stimulus onto the retina. The retina is a layer of

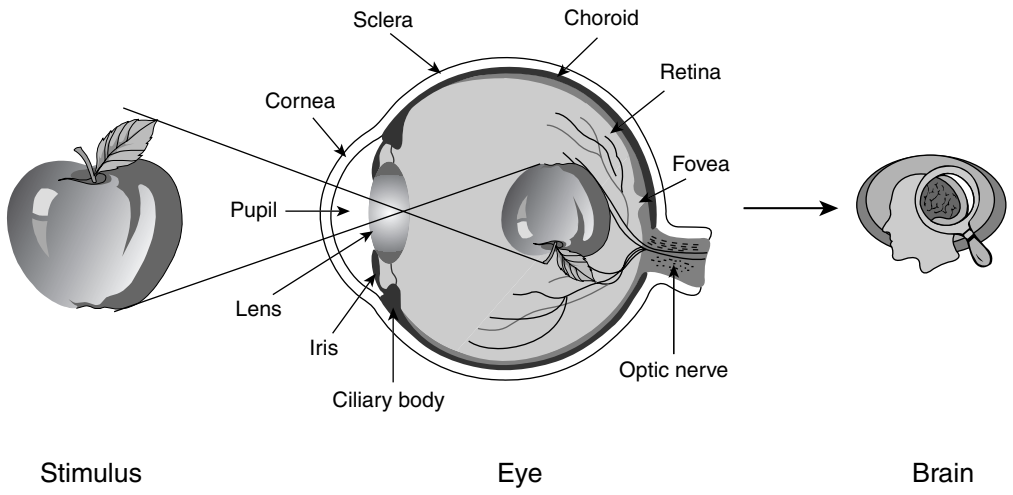


Figure 4.1 Early steps in the recognition of a stimulus object

photoreceptor cells that lines the inside of the back portion of the eyeball. The retina performs some preliminary processing of the image, after which information about the stimulus is passed posteriorly along a set of pathways toward visual areas in the brain. It is in these areas where the bulk of pattern recognition and other perceptual processing takes place. Figure 4.1 shows an overview of these steps.

According to the template matching theory of pattern recognition, an image generated by an external stimulus is matched to an internal mental representation of the stimulus, called a **template**. The degree of overlap between the image and the template is then computed. This overlap is a measure of how similar the two are to each other. A high degree of overlap will produce recognition of the object.

Evaluating Template Matching Theory

The problem with this approach is that there is a wide range of variation and possibility for any individual stimulus. Take the simple case of the letter "A." It can vary in terms of its location in one's visual field, its size, shape, and orientation. For template matching to work, there must be a separate template for each of these possibilities. Imagine how many templates for just one letter of the alphabet would be required. We would need a template for an "A" that appears to the left or the right side of the visual field, other templates for big and small "A's," others for "A's" written in script or appearing in different

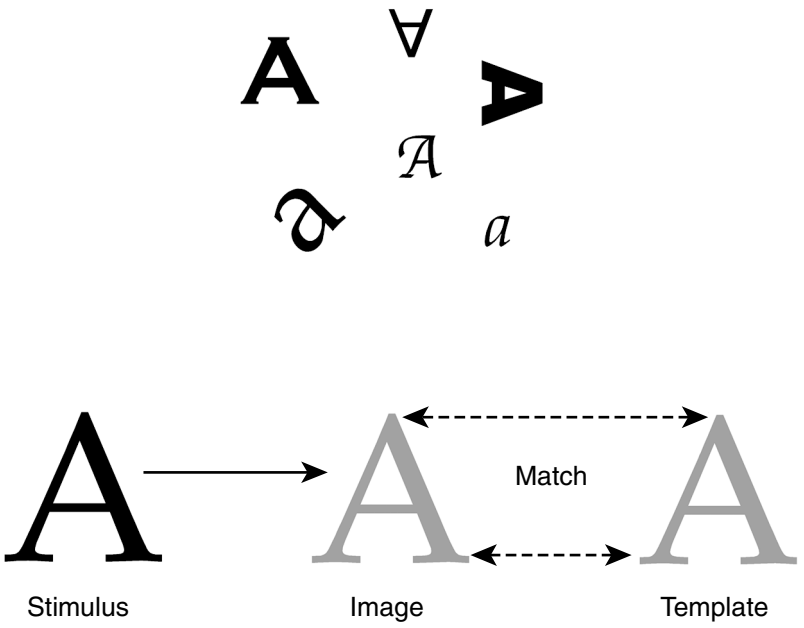


Figure 4.2 The top panel provides various depictions of the letter "A," all recognizable. The bottom panel shows the sequence of events in the template matching model

fonts, yet others for "A's" that are horizontally placed or upside-down. The number of possible templates for an individual object multiplied by the number of possible objects one could recognize becomes so large that it would become impractical to store templates or to use them effectively in memory. Figure 4.2 shows the template matching process and some examples of the ways in which a stimulus can vary.

For this reason, template matching is quickly dismissed as a theory of the way in which humans recognize patterns. It is considered a "straw man" theory, in that it is insubstantial. As a theory, though, it helps us to conceptualize what is required of and what gets in the way of a pattern recognition process. Several computer versions of template matching have been successfully implemented. One is designed to read checking account numbers off the bottoms of checks. In these cases, the procedure works because variability in the stimulus has been reduced. The sizes, styles, and the numbers of the digits that must be matched are relatively unvarying. These systems would fail under real world conditions, in which stimulus variability is much greater.

Feature Detection Theory

With feature detection models, an image of a stimulus, rather than being matched in its entirety to a template, is broken down into its component features. A feature is a part or subset of an object. The idea is that each different combination of features uniquely specifies a different object. The upright letter "A" can be specified by the fact that it has a short horizontal line as one of its features and two longer diagonal lines as additional features. It can be distinguished from the upright letter "B," which has a long vertical line and two short loops as features.

The best-known feature detection model is called the **pandemonium model** (Selfridge, 1959; Norman, 1972). It gets its name from little mental "demons" that represent processing units. These demons "shout" during the recognition process (creating a pandemonium). Figure 4.3 depicts a representation of the model in terms of how it might be used to identify the letter "R." First, the stimulus, a letter "R," is presented. It is represented by an *image demon,* which preserves the overall appearance of the letter. In the next step there are *feature demons.* There is one feature demon for each possible stimulus feature. A feature demon shouts if he sees his own feature in the image. The upright letter "R," in this example, can be thought of as having a vertical line, a diagonal line, and a loop, and so has three feature demons. These feature demons would shout out in the presence of their own features. In the next step there are *cognitive demons,* one for each possible letter. If they hear any of their corresponding features, they too will shout. The cognitive demon with the most features shouts the loudest. Finally, there is a *decision demon.* It listens to the cognitive demons. Whichever one shouts the loudest is chosen by the cognitive demon as the recognized letter.

Evaluating Feature Detection Theory

The pandemonium model of pattern recognition represents a significant improvement over template matching. It doesn't require an extensive number of templates, and will have only as many feature demons as there are features and cognitive demons as there are letters or other objects. The model can also explain the types of mistakes that people make during recognition. Often, an individual will confuse two visually presented letters that have features in common. Continuing with our example, a person might mistake the letter "R" for the letter "B," because both have a vertical line and a loop. The decision demon might accidentally choose the cognitive demon that represents the "B" because it will be shouting almost as loud as the "R" cognitive demon.

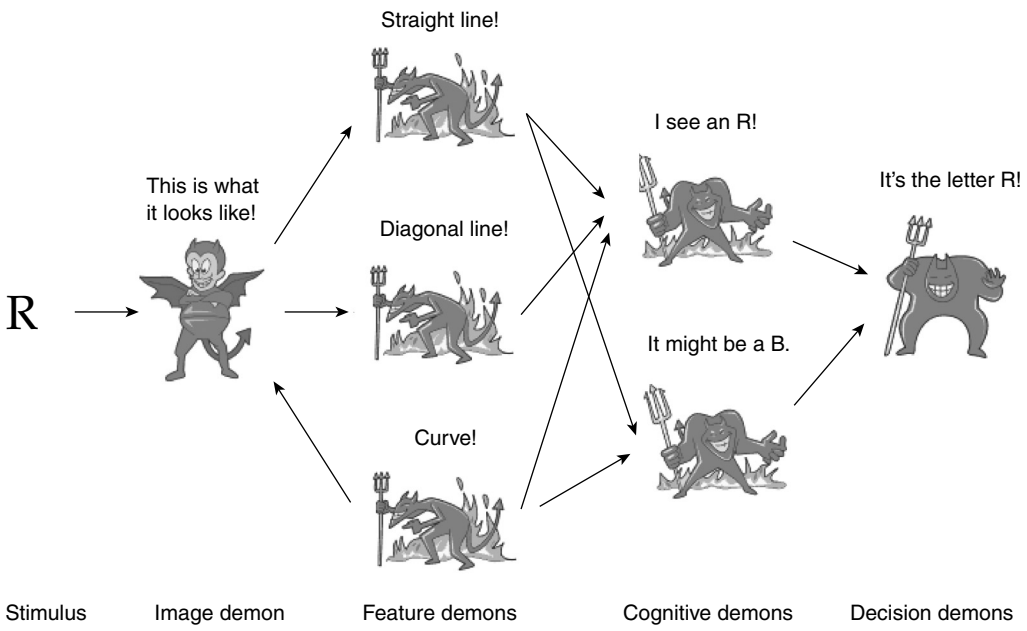


Figure 4.3 The pandemonium model of pattern recognition. Different "demons" perform different steps in the recognition of the letter R

Another reason to like the pandemonium model comes from neurophysiology. Evidence from this field demonstrates that neurons in the visual system act as feature detectors. Single-cell recordings of neurons in the primary visual cortex of cats show that these cells respond selectively to different features, such as a line of a given length and orientation (Hubel & Wiesel, 1962).

But feature detection models are not without their share of difficulties. For one thing, there is no good definition of what a feature is. For the letter "R," the vertical and diagonal lines could together form the single feature of an angle, as opposed to their being separate features. Feature detection models are also bottom-up, or **data-driven processes**. This means they are driven entirely by the physical characteristics of the stimulus itself and fail to take into account the larger context or meaning. In real world recognition scenarios, objects appear in varied contexts. The surrounding conditions often yield clues that can aid identification. In Figure 4.4, the central figure, when viewed alone, is ambiguous. The features alone are not enough to produce recognition. When

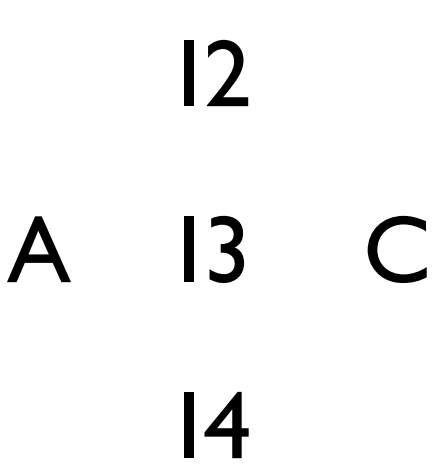


Figure 4.4 Effects of context in perception. What is the central figure?

this figure is grouped with surrounding figures, in this instance letters or numbers, it then takes on a new meaning—it appears either as a "B" or a "13."

Additional evidence for the role of context in recognition comes from what is called the word superiority effect (Reicher, 1969). Here, the time it takes to recognize the letter "A" in "CAT" is shorter than the time it takes to recognize it alone or as part of a nonsense word. This suggests that the overall word is processed before any of the letters. Because the letter "A" appears in the word CAT, the presence of the entire word facilitates recognition of the individual letter. Both the example of the ambiguous figure given above and the word superiority effect are examples of top-down or **conceptually driven processes**, in which context and higher-level knowledge aid recognition.

A Computational Theory of Vision

In 1982 the renowned cognitive scientist David Marr wrote a book that described his influential theory of pattern recognition. He adopts a computational approach to vision. In this approach, the visual system is treated as a computer. The "computer" receives an image of an object and, after performing several algorithmic processing steps, accomplishes recognition of the object. Each major processing step produces a more fully articulated version of the object—what Marr calls "sketches."

In the first stage the image of an object is presented to the retina. This image is "analyzed" in terms of the intensity values, or areas of light and dark, that

make up the image. Adjacent regions of sharp contrast in the image (that go from light to dark or vice versa) indicate the presence of edges and contours. The edges and contours in turn determine the basic features of the object. In his scheme these features include line segments and circular shapes. The result is a **raw primal sketch** of the image.

The raw primal sketch then undergoes further processing. Features that are similar in size and orientation get grouped, in much the same way parts are grouped into wholes according to the Gestalt principles discussed earlier. The groups of features are then processed again to produce a representation of the object that includes its surfaces and layout. Marr calls this the **2 1/2-D sketch**.

This image is next transformed into a complete three-dimensional representation. In this **3-D sketch**, the resultant object parts are linked by axes of symmetry and elongation (see Figure 4.5). A symmetry axis is a line that divides an object into mirror-image halves. An elongation axis is the line defining the direction along which the main bulk or mass of a shape is distributed. These axes serve to hierarchically organize the parts into a coherent, large-scale object that can be recognized. According to Marr, it is this final three-dimensional construct that we perceive and are consciously aware of.

Marr's theory is one solution to the **object constancy** problem in perception. Simply put, stimulus objects as they are represented on the retina are rarely viewed in the same way again. As we discussed in the template-matching model, an object may appear in different locations, have different orientations and/or different shapes, to name just a few possible types of variation. Any pattern recognition mechanism must therefore produce a description of the object that is impervious to these kinds of changes. In Marr's model, the 2 1/2-D sketch contains a **viewer-centered description**. This is one that is particular to the viewer's point of view. Because the viewer's location, perspective, orientation, and so on, are subject to change, object recognition under these conditions is difficult. But Marr's 3-D sketch has an **object-centered description**. The object's parts are described relative to one another and are linked on the basis of shared properties and axes. The relationships between parts thus remain intact across object transformations, enabling recognition under a broad variety of conditions.

Evaluating the Computational Theory of Pattern Recognition

It is important to point out that Marr's theory is not a statement of how the visual system actually recognizes, but how it might go about doing so. Although it is based on existing experimental and neurophysiological evidence, Marr's theory is more a speculation on the processes required of any pattern recognition mechanism, human or machine. We should also emphasize that

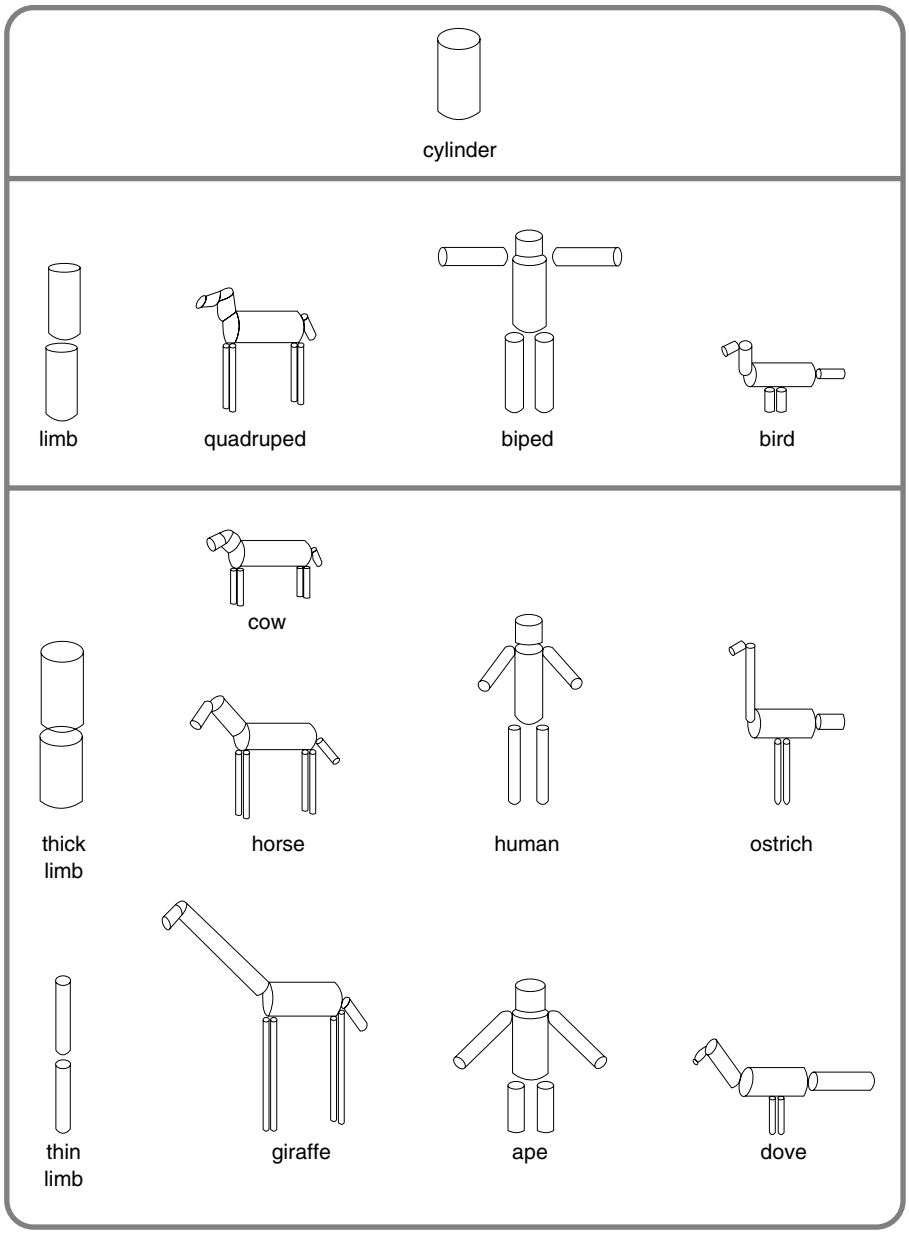


Figure 4.5 Object representations in Marr's 3-D sketch

Source: Marr, D., & Nishihara, H.K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London*, B200, pp. 269-294. Reprinted with permission.

Marr's theory involves feature detection, as the image is first broken down into component features that are then reassembled to form a complete holistic object.

Feature Integration Theory

Another theory that entails feature extraction and recombination goes by the name of feature integration theory (Treisman & Gelade, 1980). This theory is unique, because it brings to light the special role that attention plays in pattern recognition. It is based on a rather extensive set of cognitive psychology experiments, some of which we will describe below. Feature integration, like Marr's computational view, is also a stage theory. An object's basic features are identified by way of the image in the **preattentive stage**. These features can include color, motion, orientation, and curvature. Then, the features are combined during a **focused attention stage**. Following this, the object is recognized.

Let us examine more closely how features are identified and combined in these two stages. In a series of clever experiments, Treisman asked participants to identify a target item located in a visual field that was filled with nontarget items, or distractors (Treisman & Gelade, 1980). This task has since become known as **visual search**. With each trial, there is always one target, but the number of distractors can vary. For example, in one trial the target might be the single letter "T," hidden among five "S's." In another trial the "T" might be hidden among ten "S's." In each trial the time it takes the participant to locate the target is recorded. A plot that shows the time needed to identify the target as a function of the number of distractors is then made. Figure 4.6 shows a display and plot of the search function.

For the sort of task described above—locating a "T" among "S's"—the search function is flat. The amount of time it takes to locate the target is small and constant, regardless of the number of distractors. Having ten distractors does not make locating the target more difficult than having only five. Participants in the experiment report that, under these circumstances, the target seems to "pop-out" at them, so the phenomenon has since been identified as perceptual **pop-out**. Treisman argues that pop-out is preattentive. A preattentive process is one that happens automatically and effortlessly. In vision preattentive processes that are part of the processing of an image happen very quickly—usually within the first 100 milliseconds. The target item segregates out from the background of distractors and draws attention to itself before participants even know what is happening. The participants have no voluntary control over the process and don't need to pro-actively search through the display to locate the target. Treisman calls this **parallel search**, because the target

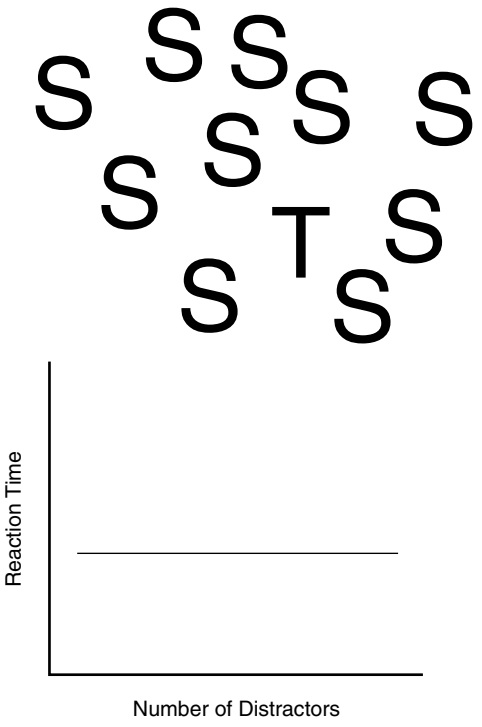


Figure 4.6 Location of the target letter in the top panel is an instance of preattentive and parallel search. The target seems to "pop-out" effortlessly. The bottom panel shows an idealized version of the flat search function for this type of task

seems to be scanned in parallel across all the display items at once. Parallel search is characterized by a flat search function. It occurs when the target differs from the distractors along a single perceptual dimension. In this example, the relevant dimension is curvature. The "T" is composed of straight lines, whereas the "S's" are curved. Pop-out also happens when the target and the distractor differ in color, motion, or brightness, suggesting that these dimensions constitute a basic set of visual features.

Now imagine a second type of visual search (see Figure 4.7). This time your job is to locate a blue "T." It is hidden among red "T's," and red and blue "S's." The search function in this case shows a linear increase. The greater the number of distractors, the more time it takes to identify the target. This suggests that a very different type of search strategy will be needed. Participants in this instance report having to look at each item in the display, one after the

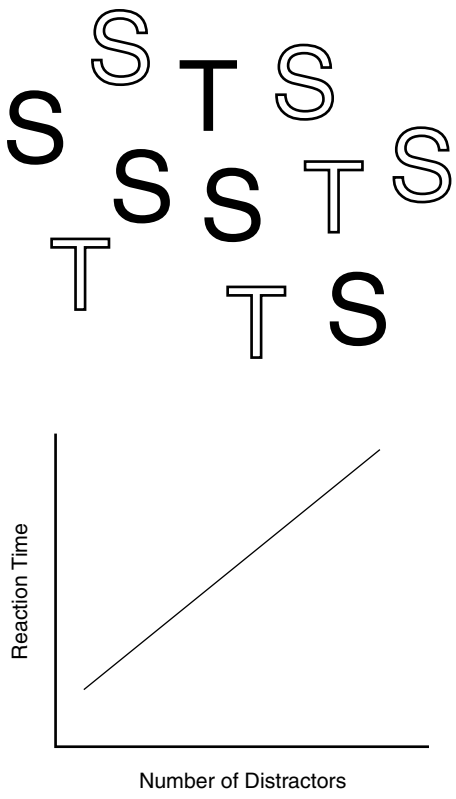


Figure 4.7 Location of the target letter in the top panel is difficult in this example of attentive and serial search. Attention must be paid to each item in the display. The search function for this task, an idealized version of which is shown in the bottom panel, is linear and increasing

other, until they find the target. For this reason Treisman calls this **serial search**. Serial search is characterized by a linear search function and appears to require focused, concentrated attention that is under voluntary control. Participants apparently need to match each item with respect to two dimensions. They must ask themselves if an item is blue and if it is shaped like a "T." If the item fails to satisfy either of these conditions, the participants must then turn their attention to the next item and do the same thing, over and over, until the target is located. Serial search is thus necessary when the target differs from the distractors along two perceptual dimensions, such as shape and color. Treisman says the job of voluntary attention, in this situation, is to "glue" together the different features. This attentional glue binds

features together into coherent objects—a necessary step before recognition can take place.

So let's review what we know of Treisman's theory up to this point. In the preattentive stage, an object's features are identified. This process is fast, automatic, and effortless. In other words, it does not require attention and happens before attentional processes can even be brought to bear. A blue "T" in this stage is immediately broken down into its basic components of the color blue and its particular line segment orientation. In the subsequent focused attention stage, concentrated attention is required to glue these different features together in one place so that they form an object representation. The color and line segments are effectively linked together by attention in one location in the visual field. Only the combination of such features can tell us what an object really is and whether it differs from other objects with which it shares some of those features.

You may encounter a real life example of visual search as you're about to leave the shopping mall. You look out into the parking lot to try to find your car, a red Honda Civic. If your car is the only red car or the only Civic in the lot, it will pop out, and you will locate it immediately. If, however, there are other red cars and/or other Civics of different colors, you will have to focus attention on each car in the lot before you will be able to locate yours.

Evaluating Feature Integration Theory

Feature integration theory has been widely accepted in the community of perceptual researchers. Investigators have conducted experiments that have generated numerous replications and extensions of its basic findings. An issue arising out of the theory, however, is exactly how and where in the visual system features get bound together. This is not a trivial problem, because, as we discuss elsewhere, the visual system uses a "divide and conquer" strategy. It breaks up an image into separate streams that correspond to object identity and location. Distinct parts of the brain process distinct aspects of a stimulus, such as its color, form, and motion. So if focused attention is the glue that binds such features, where does it all come together and how does the brain do it? This is referred to as the binding problem in vision. An intriguing solution to the binding problem is that neurons that represent separate features may synchronize their firing rates. This synchronization may then serve to link the features. Focused attention, perhaps mediated by other parts of the brain, could be the driving force behind this process. See the In Depth section at the end of the neuroscience chapter for a more complete description of this point of view.

Theories of Attention

We have already alluded to attention, having described the role it plays in the extraction and formation of features in pattern recognition. In this section we will examine the nature of attention and describe several theories of how it is used in a wider cognitive context.

Attention is defined as concentrated mental activity. In general, we can think of attention as a form of mental activity or energy that is distributed among alternative information sources. Informational sources can be stimuli from the outside world or thoughts or other forms of internal mental content. When attention is allocated to one source, that source comes into conscious awareness and the processing of that information is usually facilitated. From a cognitive perspective attention can be considered as having two sets of opposing characteristics. It is selective but divisible, as well as shiftable but sustainable. Let's discuss each of these.

By **selective**, we mean that attention can be focused onto one source of information and away from another. For example, in class you can decide to listen to the sound of the instructor's voice or to some competing sound, such as the conversation of two students in the row behind you. At the same time attention is **divisible**. It can be split or divided among several alternative information sources. You could, with some effort, listen to both the voices of the instructor and the students. In this situation, the ability to process information from the two sources would be significantly reduced. Generally, the greater the number of sources among which attention is divided, the less attention there is to devote to any single source and the worse is the individual performance with respect to each.

Implicit in the notion of selectibility is **shiftability**. One can selectively attend to information source A while ignoring B, then switch back to B and ignore A. Or, one could attend source A, then B, then C, and so on. The point is that attention can be repeatedly shifted among different sources. This shifting can be voluntary or involuntary. Distraction is an example of the involuntary shifting of attention to some irrelevant or undesirable piece of information. The converse of shifting is **sustainability**. Here, attention is focused exclusively on one source and sustained in this way over time. We can think of sustained attention as extended focus or concentration.

There are two general classes of theories that attempt to explain attention. **Bottleneck theories** describe why it is that of all the information that is presented to us, only a small portion of it actually gets through. These theories attempt to explain the apparent bottleneck effect or the narrowing down of the information

that reaches conscious awareness. Bottleneck theories are inherently theories of selective attention because they describe how some information is selected for processing as the rest gets discarded. Broadbent's filter model, Treisman's attenuation model, the Deutsch-Norman memory selection model, and the multimode model are all bottleneck theories. **Capacity theories** on the other hand are essentially theories of divided attention. They conceptualize attention as a limited resource that must be spread around different informational sources. Kahneman's capacity model is an example of a capacity theory.

Broadbent's Filter Model

The British researcher Donald Broadbent did some of the earliest pioneering work in cognitive psychology. In his 1958 book *Perception and Communication*, he proposed a theory to account for why we can't follow two streams of information coming in simultaneously via the two ears. Before we get to the details of his filter model, we need to describe some of the experimental evidence that led to it.

The **dichotic listening task** has been widely used to study selective attention. In this paradigm, a participant wearing headphones listens to two different messages being played, one over each ear. For instance, a male voice may be reciting a list of words into the left ear as a female voice recites a different list of words into the right ear. The participant's task is to pay attention to, or shadow, one of the voices and to ignore the other. The messages or information sources presented to the participant from both ears are in this context referred to as channels. Afterward, the participant is asked to recall what he or she has heard, for both the shadowed and unattended channels.

Early dichotic listening studies showed that participants were fairly accurate in recalling content from the shadowed ear, but quite poor for recalling that coming from the unattended ear (Cherry, 1953). To Broadbent, this suggested the presence of a filter that could block out messages that were being relatively ignored, allowing only attended information to enter awareness. Figure 4.8 shows a diagram of Broadbent's filter model. Information from both channels first arrives in sensory memory. This is a place where incoming sensory information is briefly held before it is processed. Next, a filter is applied to this information. The filter selects certain portions of the information, that coming from the shadowed channel in this case, and allows it to pass through. Information coming from the unattended channel is blocked.

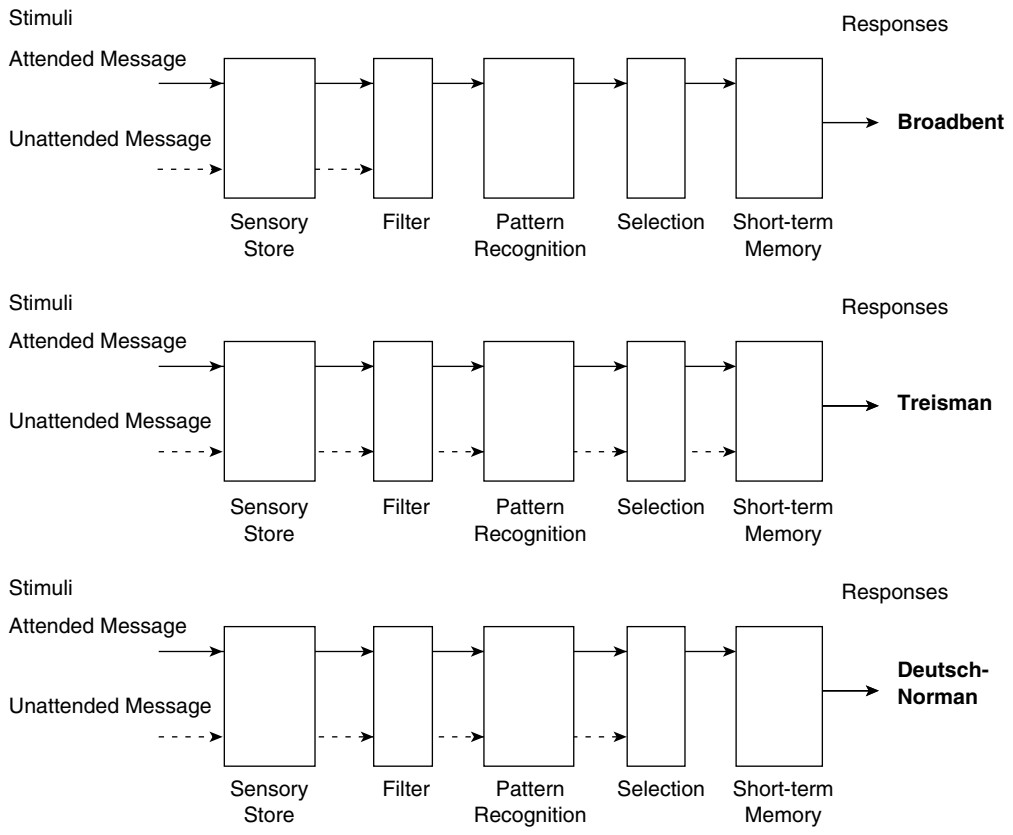


Figure 4.8 Information processing models of selective attention. The solid arrows indicate information from the attended message. The dashed arrows indicate information from the unattended message

Source: Adapted from Reed, S. K. (2000). *Cognition* (5th ed.). Belmont, CA: Wadsworth Thompson.

The selection in Broadbent's model is based on the physical characteristics of the stimulus, such as location (left or right ear), pitch, and loudness. This conclusion was based on the fact that one could shadow messages easily using these as criteria. In the next step, whatever information gets past the filter then undergoes pattern recognition. Following this step, information travels past any selection mechanism to a short-term memory store, where it is held for a longer period of time and made available for subsequent processing and response. Notice that selection in this model is performed entirely by the filter and not by any selection mechanism that comes later. We describe this

selection mechanism only to show how this model compares to others. Broadbent's model is referred to as an **early selection model** because the filter screens out information before it can be recognized.

Evaluating the Filter Model

Almost from the get-go, it was realized that there were some problems with this version of the theory. One problem is summed up in what is one of the better names for a psychological phenomenon, the so-called **cocktail party effect**. Imagine that you are attending a party and chatting with several people. Somebody across the room then mentions your name. To your surprise, you realize that you hear it, even though you were completely engrossed in the immediate conversation. Researchers have empirically validated this intuitive finding (Moray, 1959; Wood & Cowan, 1995). The observed cocktail party effect and other studies have demonstrated that some information gets through the unattended channel. The filter does not block words of great personal relevance, such as your name and words associated with danger, for example, "fire." Broadbent's model has the filter completely blocking all information coming from the unattended channel. Clearly, his view requires some revision.

Treisman's Attenuation Model

Attenuation theory is similar to filter theory but with one exception (Treisman, 1964). Instead of completely blocking out the unattended message, the filter attenuates or weakens it. Message attenuation is like having water running through two faucets, where each of the faucets constitutes a different information channel. The faucet corresponding to the shadowed message would be open all the way, allowing water to pour through. The faucet corresponding to the unattended channel would be only partially open, allowing just a small amount of water through. The addition of Treisman's model to Figure 4.8 would have the unattended message passing through all of the processing stages, but in a weakened form.

Treisman explains that different words have different chances of making it through the unattended channel, due to a threshold effect. A **threshold** is the minimum amount of activation required to produce conscious awareness of a stimulus. Stimuli with associated low thresholds easily make their way into awareness. Those with associated high thresholds do not. A word's meaning determines its threshold. Important words and those with personal relevance, such as your name, have a lower threshold for recognition and make it past the

filter. Less important words, for example, "chair," have higher thresholds and are filtered out.

The Deutsch-Norman Memory Selection Model

The cocktail party effect shows that meaning, or semantics, is another criterion according to which information is selected. This runs contrary to early filter models that posit physical characteristics such as loudness or pitch as the primary basis for selection. Deutsch and Deutsch (1963) and Norman (1968) therefore proposed a second selection process, one that is based on semantic characteristics. This selection happens later on in processing. Their model is thus an example of a **late selection model**. The first stages of processing in this model are the same as those of the other models. Information from the sensory store is filtered on the basis of physical characteristics and then recognized. However, before being passed into short-term memory, it goes through a secondary selection mechanism. This mechanism selects information on the basis of semantic characteristics or message content. The selected items end up in short-term memory and awareness. Those not selected never reach awareness. The Deutsch-Norman model shown in Figure 4.8 depicts information from two channels reaching the selection mechanism, which makes the choice as to which information gets through.

The Multimode Model of Attention

So where does selection take place? A more current view is that selection can occur early or late. Michael Posner and Charles Snyder (1975) advocate the idea of a "moveable filter," one that can operate at various points in processing according to the observer's needs. Data from another study support what seemed to some to be the case all along, that selection can be based on physical or semantic properties (Johnston & Heinz, 1978). They found, however, that semantic selection imposes a greater cost, that is, it requires greater attentional resources than physical selection. The model that arises from the view that selection can be based on multiple modes, for example, the physical and the semantic, is called the **multimode model of attention**.

There is also neurophysiological evidence to support the multimode model. Hillyard, Hink, Schwent & Picton (1973) used an event-related potential (ERP) technique to investigate when selection occurs. The ERP records brain electrical activity that happens in response to the presentation of a stimulus.

According to their design, the researchers had participants attend to a stimulus, say, coming from the left ear, and compared the activity observed under that condition with observable activity related to an identical unattended target, say, coming from the right ear. They observed a very rapid change in electrical activity only 80 milliseconds after the stimulus was presented, which suggested that selection was happening early on in response to perceptual characteristics. However, other ERP studies showed a change in electrical activity that happened approximately 300 milliseconds after stimulus onset, which suggested a late selection based on task relevance and, perhaps, semantic features (Luck & Hillyard, 1994). To summarize, it appears that attentional selection is more flexible than was first assumed. It can happen early or late, and can be based on more than just a single mode.

Kahneman's Capacity Model of Attention

Capacity models describe attention as a resource. We can think of this resource as the amount of mental effort or energy required to perform a task. Like many other resources, attention exists in limited supply. At any given time, we only have so much of it that is available for use. Daniel Kahneman, in his 1973 book *Attention and Effort,* outlined a detailed capacity model of attention. In it, he described the factors that affect available attentional capacity as well as the factors that determine how this capacity gets utilized. Because we are talking about the distribution of attention to different sources, capacity models are really models that describe the division, rather than the selection, of attention.

Figure 4.9 shows Kahneman's model. Let's walk our way through it. The rectangular box at the top represents the pool of available attention. A person's arousal level is one thing that can affect this capacity. **Arousal** refers to physiological activation and is reflected in values such as cardiac and respiratory rates. Moderate levels of arousal are assumed to produce the greatest amount of available capacity. This follows from the Yerkes-Dodson law (1908), which states that there is an inverted U-shaped performance function for arousal. Performance is poor for low and high levels of arousal, wherein individuals are presumably too tired or anxious, but optimal at intermediate levels.

Determinants of arousal include such things as the amount of sleep a person has had or the amount of food he or she has consumed. If a person has not slept well or has not eaten recently, arousal, and therefore capacity, are assumed to be low. Manifestations of arousal can include such overt behaviors as finger tapping or other nervous habits that result from excessive arousal.

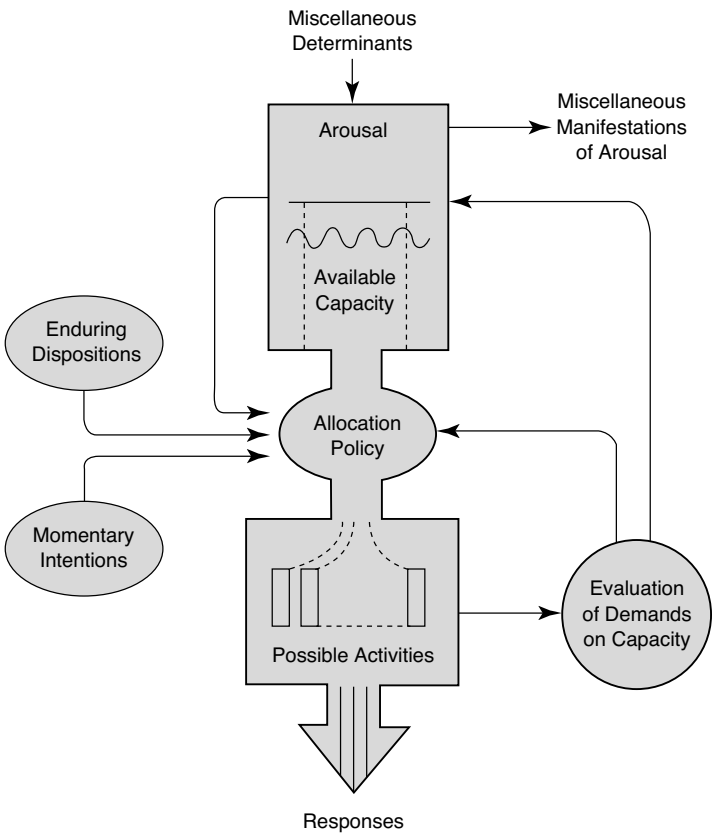


Figure 4.9 Kahneman's capacity model of attention

Source: Kahneman, D, *Attention and Effort, 1st edition*, © 1973. Reprinted by permission of Pearson Education, Inc., Upper Saddle River, NJ.

Notice too that the circle that depicts the evaluation of demands on capacity can itself affect capacity. If there is a greater demand for attention, there may be increased arousal. Knowing that you have a final examination early tomorrow morning might increase the attention you give to studying the night before.

Attentional capacity must now be allocated to the cognitive activities that need it. How is this determined? The ellipse in the center of Figure 4.9 represents the allocation policy. We can think of it as an executive or decision-maker who determines how much attention will be given to a variety of tasks. It is influenced by several factors. First among these are enduring dispositions. An **enduring disposition** is an automatic influence on where attention gets

directed. A new event or a sudden movement are things that automatically draw attention. Momentary intentions also influence allocation. A **momentary intention** is a conscious decision to pay attention to something. Momentary intentions reflect our specific goals at a given moment in time. A mother driving a car might redirect her attention to children fighting in the back seat, but she would quickly have to redirect her attention back to her primary goal of driving. Notice that the evaluation of demands on capacity and available capacity also have arrows pointing to allocation, indicating that they too exert an influence. The amount of available arousal and the urgency of a task both have an effect on the amount of attention that will go to a particular task.

The rectangular box at the bottom of Figure 4.9 contains a representation of possible activities. These are all the cognitive processes or tasks that need attention. The allocation policy that gives attention to them is flexible. For example, equal amounts of attention could be given to all the activities or all available attention could be given to just one activity. Those activities with the greatest demand will be given the most attention. In the driving example given above, the driver's attention was momentarily taken away from driving, the most demanding activity, and given to the children in the back seat, and was then switched back again.

Evaluating the Capacity Model of Attention

Stephen Reed, in his book *Cognition* (2000), makes some summary comments on attentional theories. He notes that capacity models such as Kahneman's are not designed to replace selection models, but rather to supplement them. Both types of model predict that it will be difficult for an individual to pay attention to two things at once (but for different reasons). According to selection theories, it is because a bottleneck develops, prohibiting the entry of two packages of information at the same time. According to capacity theories, it is because the demands of the two tasks exceed available capacity. Capacity models give us a good overview of the many influences and interdependencies that are involved in the allocating of attention. Selection models provide us with a better feel for the discrete stages of information flow during attentional processing.

Evaluating the Model-Building Approach

Our discussion of perceptual and attentional theories so far has demonstrated several things. First, that even within a discipline there is no single widely agreed-upon model that explains how a given cognitive process operates. In the

case of pattern recognition, we summarized five major models of how it might take place. Second, although these models may vary widely, each one captures some unique aspect of the process better than the others. The feature detection model is better suited to explain errors in recognition, whereas the computational approach can best account for object constancy. Third, and most important, we learn more about a process by evaluating all of the theories and comparing competing theories. The problems posed by the template-matching model show that any model of pattern recognition must extract some invariant feature from the image. That is in fact what is postulated in one form or another by the remaining theories. This cross-theory analysis may not tell us what the features actually are—we see they differ slightly from model to model—but it does tell us that the process of feature extraction is a key element in the process.

Another major point worth mentioning here is that there is cross-fertilization between model formation and experimention. In the case of attention, we have seen a constant tug and pull between the models and the experimental evidence. This dynamism led to the formation of better and more improved models. Broadbent's filter theory showed us the basics of how a selective attentional system might operate. Later research, however, proved that information could get through the unattended channel. This led to the formation of Treisman's attenuation model, which by itself could not adequately explain selection that is based on semantic characteristics. This in turn led to the development of the Deutsch-Norman memory selection model. The results of later experiments necessitated the introduction of multimode models and the idea of a moveable filter. Drawing on this short history, it becomes obvious that we should not put too much faith in one model. Instead, we must realize that each model is a point along a developmental continuum and also encourage researchers to modify their models based on current research.

In Depth: Biederman's Recognition-by-Components Theory of Pattern Recognition

In the preceding section on pattern recognition, we saw that there are many ways to recognize an object. A problem inherent in all of these procedures is to extract from the stimulus input aspects that are invariant with respect to viewing angle, lighting, and so on. The template matching solution was to store multiple copies of the object. The feature detection and integration solutions were to extract certain invariant features—aspects of the object that remain constant. Marr's solution was to generate an object-centered description.

The psychologist Irving Biederman (1987) also proposes a feature extraction account for the constancy problem; his approach is interesting because it

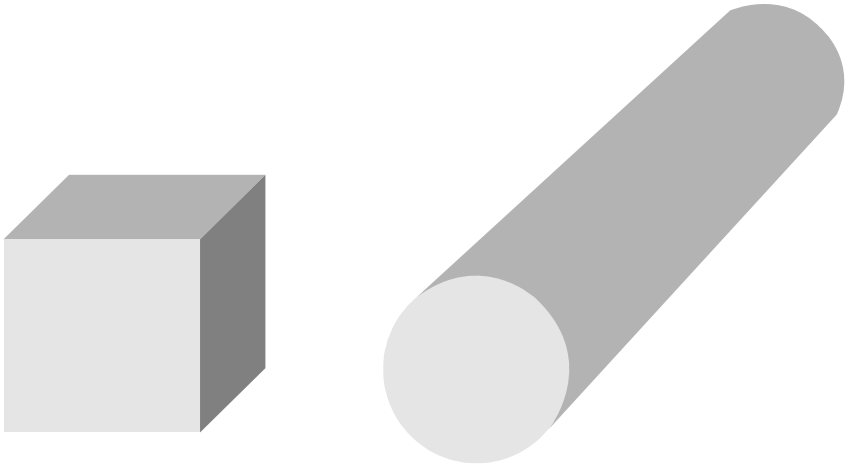


Figure 4.10 Cubes and cylinders can be considered examples of geons. These can then be assembled to form more complex objects

specifies how object constancy may be achieved for three-dimensional object representations. In his theory features gathered from a stimulus are recombined to form an object representation. These features he calls **geons**. A geon has a basic volumetric shape, for example, a cube or a cylinder. Altogether there are 36 of them. Figure 4.10 shows two examples of individual geons.

Geons are said to have three basic properties. The first is view invariance. This means geons can be identified when viewed from many different perspectives. The second is discriminability: one can tell one geon apart from another no matter what the viewing angle. The fact that geons have these characteristics supports the notion that they may be used as features in human pattern recognition, since they can account for the view-invariant aspect of object constancy.

The third property of geons is resistance to visual "noise." Geons can be perceived even when many of the contours that make them up are obscured. In one study, Biederman (1987) showed observers line drawings of objects that were masked by noise. The noise in this case was amorphous black regions resembling patches of spilled ink. In one condition, the noise covered the object contours in a way that preserved the geons. The noise for the most part in this condition covered single stretches of contour. In a second condition the noise covered the contours such that the perception of the geons was disrupted. The noise here covered contour junctions, which are needed for the construction of the surface planes of three-dimensional shapes. Observers had no trouble recognizing the shapes in the first condition, but had great difficulty doing so in

the second. These results suggest that geons may help us to recognize objects that have undergone occlusion, a partial covering up by other objects, since in the real world occluding shapes rarely obscure junctions.

Biederman's recognition-by-components theory is thus supported because it can account for view invariance and occlusion. Another strength of the theory is the small basic feature set that consists of just 36 geons. In other pattern recognition theories, features are either unspecified or extremely numerous. Critics of the theory point out that broad classes of objects can be discriminated on the basis of geons, but that fine-level object categories cannot. For example, geons can explain why we have no problem telling the difference between an airplane and an automobile, but they have difficulty in providing an explanation for how it is we can also tell apart objects that are much more similarly-shaped, such as two similar bird species (Perrett & Oram, 1993). For these sorts of differences, we must rely on other features, such as the angle or taper of a part or perhaps color and texture.

Minds On Exercise: Silhouettes and Object Constancy

We rarely view an object from the same perspective twice. This poses a problem for pattern recognition mechanisms, which must extract invariant features—those that don't change when there is some alteration of the stimulus input. To see what a difficult job the visual system has, try the following exercise. Have one person suspend an object by a string behind a paper screen. Have another person shine a light on the object, so that only its shadow appears on the screen to the observer. The observer now has to guess what the object is. Rotate the object. Is recognition now easier? Why? Were there some perspectives from which it was easy to recognize the shape? What were they? From which perspectives was recognition most difficult? Why? For each shape, record the feature or characteristic of the object that allowed you to recognize it.

Food for Thought: Discussion Questions

1. Do you think a process model is a fair way to describe mental processes? Do you agree with its assumptions that information processing is sequential and independent? Can you think of another way of representing mental processes? If so, what type of diagram would you draw to represent it?
2. Write down all the letters of the English alphabet. Compare them and determine which letters would be most easily confused with others. What features

do these letters have in common? Devise a new set of features that would allow you to tell them apart.

3. Would it be easier to recognize your mother in your house or in a gas station? What role does the context play in facilitating or interfering with recognition? Is this an example of a data-driven or a conceptually-driven process?
4. What sort of things are good at causing distraction? Are there some stimuli that are impossible to ignore? Conversely, name those things that don't divert attention and are easy to ignore.
5. How many things can we pay attention to at one time? Is it only one? Is it greater than one?
6. Do animals possess attention? Can they control their attention in the same way humans do? In what ways can humans control their attention that animals cannot?

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution.* New York: Basic Books.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Matlin, M. W., & Foley, H. J. (1997). *Sensation and perception* (4th ed.). Boston: Allyn & Bacon.

Pashler, H. (1998). *Attention.* East Sussex, UK: Psychology Press.

Reisberg, D. (2001). *Cognition: Exploring the science of the mind* (2nd ed.). New York: Norton.

5

The Cognitive Approach II: Memory, Imagery, and Problem Solving

"It isn't so astonishing, the number of things that I can remember, as the number of things I can remember that aren't so."

—Mark Twain, 1912

Types of Memory

In this chapter we continue our discussion of cognitive theories, as we focus on three other prominent cognitive processes: memory, imagery, and problem solving. We will describe the major theoretical approaches to each as well as some of the classic experiments that led to their development. There has been an immense amount of research on these and other topics in cognitive psychology that has been conducted, and by necessity we must leave out many interesting and important findings. As in previous chapters, refer the interested reader to the works in the Suggested Readings sections if they wish to learn more of these findings.

Memory, very generally defined, is the capacity to retain information over time. Memory is of course very important to any information processing system, animal or machine, because it underlies the ability to learn. Any

system incapable of learning from its mistakes would soon perish or become obsolete in a dynamic, real-world environment. Memory allows us to store our past experiences and draw on them. In this way, we can deal with new situations that bear some similarity to old ones in the way we have in the past, and not have to solve each new problem from scratch.

Any discussion of memory must begin with the notion that there is no single type of memory. Cognitive research in this area has led to a rather consistent finding: the existence of functionally distinct memory systems. Models of memory specify how these different memory systems interact. Before we can understand this interaction we must lay out some of the characteristics and features of these individual memory systems. They include duration, which refers to how long information remains viable in a memory system; capacity, which refers to how much information the memory system can hold; and coding, which refers to the particular type of information the system contains. In the first part of this chapter we will summarize these characteristics. Following this, we will describe formal memory models.

Sensory Memory

Sensory memory is a repository for incoming sensory information. Raw, unanalyzed data that are derived from the senses are held here very briefly. The purpose of sensory memory is to maintain the representation of a stimulus long enough so that it can be recognized. Although you may have glanced at a visual scene for only a very brief time, say, 100 milliseconds, a representation of the scene is preserved in sensory memory for longer than that. This gives the information a chance to be operated upon by selection and pattern recognition mechanisms.

There are different forms of sensory memory—there is one for each of the five senses. Each of these forms has different characteristics. **Iconic memory** is a visual sensory memory. It holds a brief "snapshot" of what you have just looked at. Iconic memory has a very short duration; it lasts only about 250 to 300 milliseconds (Averbach & Sperling, 1961; Sperling, 1960). **Echoic memory** is an auditory sensory store. You can think of it as an "echo" of what you have just heard. It lasts considerably longer than iconic memory, on the order of several seconds longer (Darwin, Turvey & Crowder, 1972). The coding or representation of information in sensory memory thus varies with the modality. Iconic memory stores visual representations; echoic memory, auditory ones.

George Sperling first began investigating iconic memory in the 1960s. He would present subjects with a short display of letters of the alphabet

consisting of a 4 letter x 3 letter array. In a given trial a display might look something like the following:

R G C P

L X N F

S B J Q

There were two conditions in this study (Sperling, 1963). In the **whole-report condition**, the partipants' task was to remember as many of the letters as possible. Most persons recalled only four or five. Sperling had a hunch that they were actually remembering more than this. He suspected that the procedure was testing for not what is available in the icon (in this case the display of letters), but what we pay attention to or process after the fact. Four or five individual items might be what is remembered, but not what is actually perceived in the icon. To investigate this, he developed a **partial-report condition**. Participants, immediately after being given the letter display, heard one of three tones. A high-pitched tone cued the top row, indicating that the participants would be expected to report on that row; a medium-pitched tone, the middle row; and a low-pitched tone, the bottom row. The results then showed that subjects were able to recall all of the letters, regardless of their location. This demonstrates that all the letters were stored and available in the icon, as the observers had no way of knowing which row was going to be cued.

We learn from this research that the capacity of the visual icon, how much it can hold, is at least twelve items. Subsequent research has shown that iconic memory can in fact hold much more than this. Its capacity is essentially unlimited. Everything that can be taken in during a glance—all the information in the visual field—can be stored there. But the integrity of this information is short-lived. The image is clearly represented early on, but then fades quite rapidly. The loss of information over time in memory is known as **decay**.

Working Memory

Working memory is also sometimes known as short-term memory. Some researchers sometimes use these terms interchangeably, as will we throughout this text. As the term *short-term memory* suggests, information is briefly stored there. However, the duration of items residing in working memory is, as we will see, much longer than that of items residing in sensory memory. Working memory retains less information; its capacity is unlike the unlimited capacity of the visual icon. Working memory is limited to storing just a small number

of items. Whereas information in the different sensory stores is specific to modality, coding in working memory can be acoustic, semantic, or visual. We will have more to say about how the different types of codes are put to use when we discuss Alan Baddeley's model of working memory later in this chapter.

It is helpful to use the analogy of a workbench when thinking about working memory. A workbench is a space where one can construct something using parts and tools. For example, if one wanted to build a spice rack, he or she would need several pieces of wood, some nails, and varnish. In addition to this, he or she would need saws to cut the wood to size, a hammer, and a paintbrush. With these items in hand one could then set about to create the final product. Working memory is the mental equivalent of this workbench. It is a place where data (the parts) can be temporarily stored so that they can be operated on by cognitive processes (the tools). In this sense working memory is the site where conscious thinking takes place. It is here that you remember a phone number, figure out how to navigate your way around a new city, or solve an arithmetic problem.

In a classic study, Peterson and Peterson (1959) demonstrated the duration of items that reside in short-term memory. They presented participants with items to be remembered. A tone was sounded following the presentation of the items. In separate trials the tone was sounded at varying time intervals. Sometimes it would go off immediately; during other trials there would be a substantial delay. The participants were instructed to recall the items upon the sounding of the tone. In this condition, all participants were able to remember the items correctly, no matter the length of the delay. In a second condition, presentation of a three-digit number followed presentation of the items. Participants were instructed to count backward in threes from this number. If the number was "796," they would have to count out "793," "790," "787," and so on. A tone was sounded, again at varying intervals, cueing recall of the item. At this point the results were quite different. The ability to remember the items deteriorated rapidly as the period of delay increased. After an 18-second recall interval, the accuracy of recall had dropped to just 5%, meaning participants could only recall 5% of the items correctly. Figure 5.1 shows idealized results for both conditions of such a study.

You may have guessed what's happening here. In the first condition, participants were able to rehearse the trigrams. **Rehearsal** is the mental repetition or mental "practicing" of some to-be-learned material. It is usually manifested as implicit speech, or "talking to oneself"—what you do when you need to remember somebody's phone number but can't write it down. Rehearsal refreshes items in short-term memory; it keeps them active and prevents them

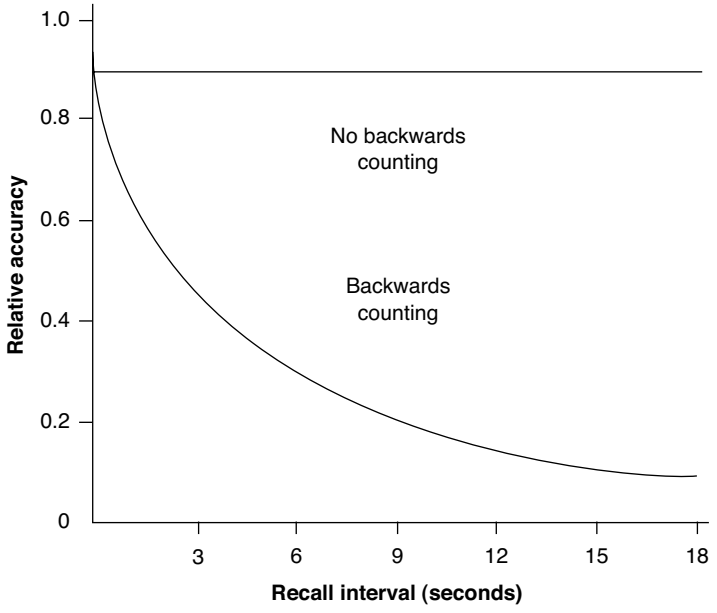


Figure 5.1 Idealized results of a memory study in which participants are asked to remember an item after a given recall interval. Having to count backwards after presentation of the item interferes with rehearsal

from decaying. In the second condition, participants were prevented from rehearsing because of the backward counting. Here the items underwent decay. Using this paradigm, Brown (1958) and Peterson and Peterson (1959) were able to determine that the duration of short-term memory was about 18 seconds. If you are interested in some of the other processes that can act on information in working memory, go to the In Depth section, where you will see that, in addition to rehearsal, items can be scanned.

Our next discussion is of capacity. How much can working memory hold? It is easy to devise an experiment that tests this question. It would involve presenting participants with a series of lists, each containing a different number of items. For example, we could present a list of four digits, then a five-digit list, a six-digit list, and so on. If there were a limit to how many digits could be held in working memory, we would see a dramatic drop off in retention as we arrived at that limit. Many early experiments of this sort demonstrated that this was indeed the case. On average, individuals can retain about seven items, give or take two, commonly denoted as 7±2. This limit has come to be called the "magical number seven" (Miller, 1956).

So short-term memory has a limited capacity. We can only hold about seven items in it at one time. But what exactly is an "item"? Is it a number? A letter? A word? To understand this, let's start with a demonstration. Glance quickly at the letters below and then close your eyes to see how many you can remember.

N F L C B S I R A M T V

Did you have a hard time? If you weren't able to recall all of the letters easily take a closer look. You may at this point notice that the letters cluster in groups of three—each group is a popular acronym. "NFL" stands for "National Football League," "CBS" for "Columbia Broadcasting System," and so on. If you had noticed these groupings, it would have been easy for you to remember all the letters. This is because you would then be remembering only four groups of three letters and not twelve individual letters. A group of meaningful items is called a chunk. The grouping of items into a single meaningful whole in short-term memory is known as **chunking**. We can increase the total amount of information that is contained in short-term memory by chunking: the bigger the chunk, the greater the capacity.

Let us turn now to the issue of coding. How is information coded or represented in working memory? Research in this area shows that information can be coded in a variety of formats. Conrad (1964) presented letter strings to participants and then asked them to perform immediate recall. He found that although the letters had been presented visually, the pattern of errors was based on the sounds of the letters. Participants would mistake an "A" for a "K," for example, or an "E" for a "P." This suggests that the letters were being converted to an **acoustic code**, one based on the sounds of the items.

Studies by Wickens (1972) and others have demonstrated the existence of an alternate code that is part of working memory. To make sense of their experiments, we must first differentiate between two types of interference. In **proactive interference**, information that is learned earlier interferes with information learned later. In **retroactive interference**, information that is learned later interferes with information learned earlier. The more closely related items are to each other with respect to meaning, the greater this interference. Wickens presented words to subjects along the lines of the tasks that were given participants in the Brown/Peterson experiments. He found that after they had studied words that belonged to a single category, there was proactive interference, which engendered a steady decline in recall over the first few trials. For example, the word "apple," learned earlier, would interfere with recall of the word "orange," learned later. But in a final trial during which the semantic categories of the words the participants were asked to remember were sometimes switched—in one condition, from fruits to flowers—there was an increase in

recall. In this instance, the word "apple," learned earlier, would interfere less with the word "tulip," as they belong to separate semantic categories. These results showed that the words were represented not acoustically, but instead on the basis of their inherent meanings—the code was a **semantic code**.

Additional studies showed that information in working memory may be represented in yet a third way. A **visual code** is one that preserves spatial characteristics. In studies of the mental rotation of objects, participants were presented with a visual pattern and were asked to compare it against another that was either a different pattern or the same pattern rotated by a certain angle. The plot of reaction time in these studies showed that subjects apparently formed a visual representation of the pattern and rotated it mentally to enable them to perform the comparisons (Shepard & Metzler, 1971; Cooper & Shepard, 1973). The representation must be visual because the greater the angular difference between the two patterns being compared, the longer it takes to respond—exactly what one would expect if a real spatial object were rotated. We will describe these experiments and the controversy that surrounds them in greater detail later in this chapter.

Long-Term Memory

Obviously, individuals are capable of remembering information for longer than just a few seconds. Many of us can recall quite vividly events from our childhoods or facts learned decades ago in school. We must therefore have another memory system, one that is capable of storing data for longer periods of time.

It turns out that, as is the case for memory in general, long-term memory has several distinct types. **Procedural memory** holds procedural knowledge. It is memory for skill, is demonstrated by doing, and arises without conscious recall. Knowing how to ride a bicycle is a good example. A person who knows how to ride a bike can demonstrate that he or she has this ability really only by actually doing it. The subconscious nature of this memory becomes evident when we acknowledge that it sometimes happens that we learn how to perform some skill, such as playing the piano, forget how, but then show improvement when we attempt to perform the action at a later date. Procedural memory also sometimes goes by the name of implicit memory.

Declarative memory contains declarative knowledge. It is memory for facts and events, is demonstrated by speaking, and arises with conscious recall. There are two types of declarative memory. Knowing that Thomas Jefferson wrote the Declaration of Independence is an example of the first type, called semantic memory. **Semantic memory** is knowledge of facts and

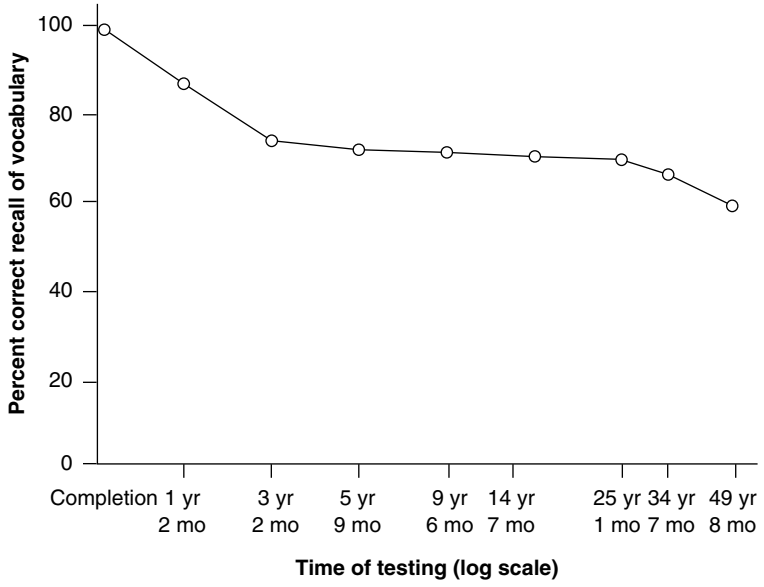


Figure 5.2 Percent correct recall for Spanish learned in school, at different time periods after original learning

Source: Adapted from Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113*, 1–29.

general knowledge of the sort learned in school. The second type is episodic. **Episodic memory** contains episodes, or personally experienced events, for example, what you did on your birthday last year. We are usually consciously aware of declarative information, which is sometimes referred to as explicit memory.

A study by Bahrick (1984) shows us how long information in semantic memory lasts. Bahrick studied a group of people who studied Spanish early in life, usually during the high school or college years. They were tested at various times throughout their lives with respect to how much they remembered from their Spanish classes. When recall for various types of information was plotted as a function of time, three distinct memory stages were discovered (see Figure 5.2). In the first stage, stretching from 3 to 6 years after the classes, there was an initial rather dramatic loss of information. In the second stage, stretching from 6 to up to 30 years after initial learning, there was no further loss—the amounts remembered stayed relatively constant. In the third stage, stretching from 30 to 35 years later, there was another loss of information, but only a slight one.

We can draw conclusions from each of the three stages. First, information in long-term storage, if not used or rehearsed, decays fairly rapidly over the first few years. Second, there is some residual amount of learned information that remains intact over a long period of time. Third, there appears to be a slight additional loss of information that occurs later in life. This loss could be due to the kind of general decline in cognitive function that accompanies aging. The basic shape of this memory function stayed the same regardless of the levels of training or the grades the participants had received. Better initial learning, as measured by these indices, resulted in an elevation of the function such that overall performance was better, but the shape of the function remained unchanged.

Now back to capacity. How much can be retained in semantic long-term memory? It has been proposed that we remember virtually everything we've ever experienced in our entire lives, but simply have difficulty recalling it. Therefore, although information may get into long-term memory and stay there without its being lost, our inability to remember it could be due to a failure in "getting it out." One researcher estimates that the average adult has about a billion bits of information in memory and a storage capacity that is perhaps one thousand to one million times greater than that (Landauer, 1986). However, we must be skeptical about such estimates, since the inability to recall an item in a memory test can be due to either retrieval failure or decay.

With respect to coding, the current belief is that information in long-term memory is represented in various formats. Implicit memories may be stored in the form of production rules—formalized if-then statements that match a sensory input to a motor output. The proposed neural locus for production rules is in the cerebellum, a part of the brain that mediates motor learning. Refer back to the introductory chapter for more on production rules. Explicit memories may be stored as networks of connected nodes, each node representing a fact or event that is linked to others. The distribution of these nodes is most likely throughout the entire cortex and other diffuse brain areas. See the network approach chapter for more on these networks.

Memory Models

We turn now to summarizing some of the most influential theories of information processing that attempt to describe memory. These theories specify interactions between all the different memory types we've seen so far. Alternatively, some of these theories describe the interactions between components of a single memory system. In each case, the interactions are conceptualized through the

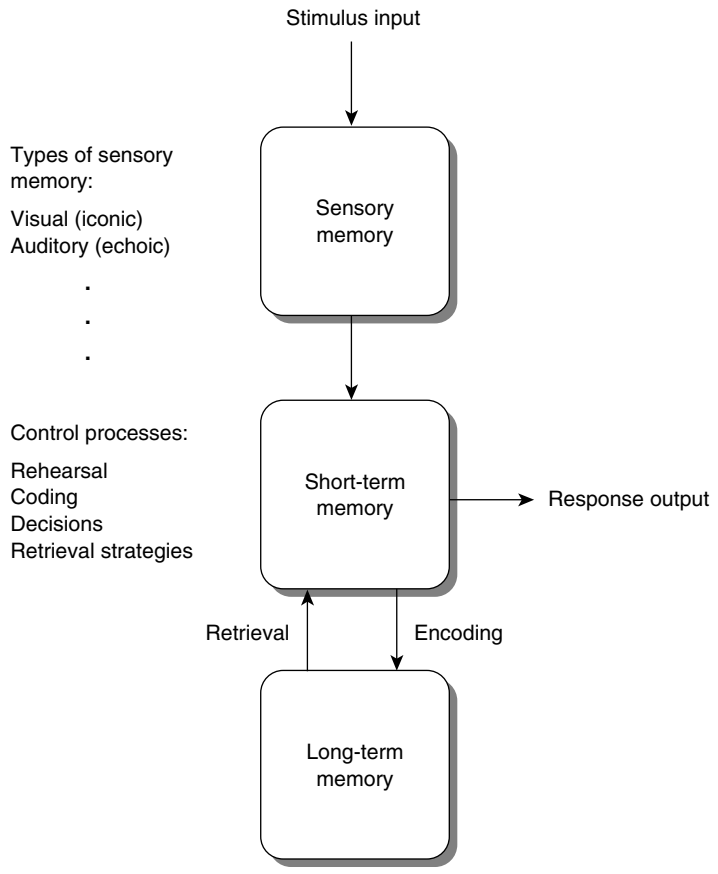


Figure 5.3 The Atkinson & Shiffrin (1968) modal memory model

use of process models that show how information is transformed and shuttled back and forth among the successive stages of processing.

The Modal Model

The **modal model** of memory was the first model to provide a general overview of how information is processed in each of the different memory types (Atkinson & Shiffrin, 1971). Figure 5.3 shows the stages of the modal model. To start, stimulus information from the outside world first enters sensory memory where, as we have seen, it is very briefly held. From there, the

information is transferred to short-term or working memory, where a wide variety of computations can be performed. These include rehearsal and recoding. Following this, to-be-remembered information is passed to long-term memory. **Encoding** is the name of the process by which information is taken into long-term memory and converted into a usable form. A related process, discussed at length in the neuroscience chapter, is consolidation. Consolidation is another process by which information is transferred from short- to long-term memory. Consolidation "strengthens" information so that it lasts longer.

Once in long-term memory, information may not be used immediately. We say that the information is in a state of **storage**. Stored information is represented, but not available for immediate use. As alluded to above, this information may sit in limbo for a very long time. When information that resides within long-term memory is needed, a retrieval process takes place. **Retrieval** is the act of accessing needed data and making it available for use. In the modal model, retrieval is shown as the transfer of the information from long-term to working memory, where it can then be operated on.

Evaluating the Modal Model

The modal model provides us with a good summary of memory function and serves as a useful starting point. However, it was developed very early in the course of cognitive psychology and fails to specify many of the nuances of memory structure and function. Upgrading the model would require the postulation of additional memory subtypes and the insertion of a number of additional processing steps. For instance, inside the long-term memory box we would need to designate distinctions for implicit and explicit memories. In addition, the modal model fails to acknowledge the large number of information processing operations that can be performed in working memory. More recent models of working memory do just this and are discussed later in this chapter.

The ACT* Model

John Anderson (1983, 1990) proposes a global model of memory function that is similar to the modal model. Anderson's model is really more than just a description of how memory works. It is considered a cognitive architecture—a concept we discuss later. However, we talk about it here in the context of memory. The layout of his design, which he calls ACT* (read "act-star," where ACT stands for Adaptive Control of Thought), is shown in Figure 5.4. A revised version, ACT-R, has also been formulated (Anderson & Lebiere,

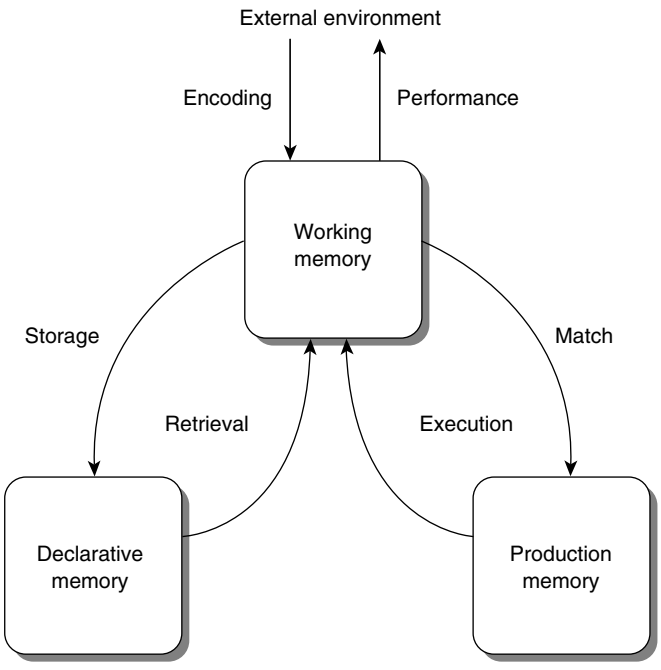


Figure 5.4 Anderson's (1983) ACT* model of memory

1998). Anderson's model has three components: working memory, declarative memory, and production memory. Declarative memory is equivalent to explicit memory and production memory is equivalent to implicit memory. The arrows in the figure indicate different kinds of processing that occur as information is exchanged between the memory systems.

Let's use two examples to illustrate how the model works. The first demonstrates the use of procedural knowledge. Imagine that you are driving in your car and you see a stop light. The stimulus of the red light is encoded into working memory. From there the stimulus is weighed against a production rule in production memory. If there is a match, the corresponding command to step on the brakes is given. This command is passed back to working memory in the step labeled execution. The command is then translated into action in the process of performance. In the model, this encoding-match-execution-performance loop is activated when one is carrying out any learned skill, whether driving, skiing, or playing the piano.

Now imagine that you are answering a factual multiple-choice question on a college exam. A representation of the question enters working memory and

is transferred to declarative memory, where semantic knowledge that is related to the question is located. This transfer step is called storage. The question then activates the answer and it is retrieved back into working memory where it can be applied to the question. Once an appropriate response is selected, you would write in the answer, this action of writing being the performance. This encoding-storage-retrieval-performance loop is activated when one is using declarative knowledge.

Evaluating the ACT Model*

The strength of Anderson's model lies in its description of separate processing loops for implicit and explicit memories. Notice that there are no arrows that connect production memory and declarative memory. These are considered two entirely distinct and independent memory stores. They do not share information or communicate with each other. You should remember that conscious awareness accompanies explicit, but not implicit, memory. The absence of any information exchange between these two systems reflects this dichotomy.

Anderson postulates an additional component of this model that represents how propositional information in declarative memory is organized. This organization takes the form of a network in which nodes represent concepts and links represent the relationships between them. He also allows for different types of nodes that stand for specific examples of concepts we've encountered, for example, a specific person we know. In this way, the network can represent episodic as well as semantic information, both of which are part of declarative memory.

The Working Memory Model

Baddeley (1986, 1992) has formulated a detailed model for the components and processes of working memory. These are shown in Figure 5.5. In Baddeley's scheme, working memory is composed of three separate units. He calls the primary unit the **executive control system**. The job of this system is to initiate and control ongoing processes. Like the captain of a large ship, this system issues orders and directs subsidiary crews. Some of its activities are reasoning, language comprehension, information transfer to long-term memory via rehearsal and chunking, and retrieval of information, also from long-term storage.

The second major component in Baddeley's model is the **articulatory loop**, sometimes called the phonological loop. This is the place where speech and sound-related information are rehearsed. Information such as a telephone

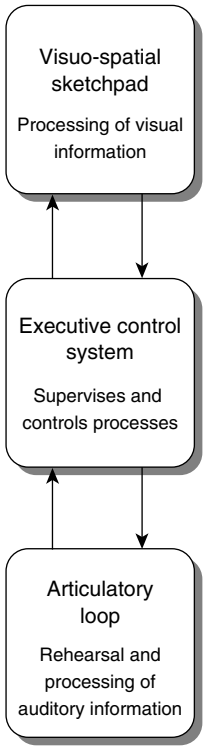


Figure 5.5 The different components of the working memory model (Baddeley, 1986)

number that you want to remember and actively maintain for immediate recall is passed to the articulatory loop to refresh it. A third unit, called the **visuo-spatial sketchpad**, is specialized for the processing of visual information. It is here that visual imagery tasks such as mental rotation or visual search are performed. If you were to form an image of a cat with your eyes shut, the image would be represented in the visuo-spatial sketchpad.

Both the loop and the sketchpad are "slave systems" to the central executive. They carry out processes such as rehearsal and image formation. Other processes, such as reasoning and language, are the responsibility of the executive. Notice also that these two slave systems are domain-specific: one is devoted to the processing of acoustic information only, the other to the processing of spatial information only. This means that each one operates on a different code; the loop, on an acoustic code, and the sketchpad, on a visual code. Tests of working memory have shown that each slave system relies on its own pool of attentional

resources (Baddeley & Hitch, 1974; Logie, Zucco & Baddeley, 1990). If the tasks that one of these systems is performing are simple and don't require much in the way of attentional resources, then working memory performance is unchanged. If one of the systems is given a demanding task, though, it either fails to complete the task or draws on the attentional resources of the executive, which results in impaired working memory performance.

Evaluating the Working Memory Model

Baddeley's working memory model is in essence a reformulation of the traditional idea of short-term memory. This idea would have us believe that short-term memory is a relatively inactive location where information is either rehearsed or it decays. Baddeley extended our concept of this memory system to encompass a self-regulating collection of processors that perform imagery, reasoning, and language tasks. Recent studies have shown that there are specific brain regions that correspond to the different components of the Baddeley model (Smith & Jonides, 1999; Smith, 2000). Verbal storage, rehearsal, and executive tasks were found to activate a number of regions in the parietal and frontal lobes of the left hemisphere. In contrast, visual processing activated regions in the occipital, parietal, and frontal lobes of the right hemisphere. Much work in the area of uncovering exactly which types of neural activity underlie the specific operations of working memory remains to be done. For example, we do not as of yet have a detailed and accurate neural processing model that can account for the performance of various mental arithmetic tasks, such as addition.

Visual Imagery

How many windows are there in your living room? What is the shape of a German shepherd's ear? In order to answer these questions, you may have resorted to the use of visual images. A **visual image** is a mental representation of an object or scene that preserves metric spatial information. Visual images are therefore isomorphic to their referents. They preserve spatial characteristics of the objects they represent. An example would include spatial extent. If it takes longer to scan what is in-between two parts of a physical object because the distance between those parts is relatively greater, then it should also take longer to scan in-between those parts in the image representation.

Isomorphisms do not imply that images have the same physical properties as the objects themselves. An imagined banana is of course not really yellow or

curved. But the representation of the banana in the image allows it to be treated as if it were a real banana. That is, it may be *mentally* picked up, peeled, or squashed. The image allows us to perform the same sorts of operations on the banana in our heads as we would on the actual banana with our hands. Although images in many ways seem to follow the same laws as real objects, they are by no means constrained by such laws. We could, if we wanted, imagine a flying banana or a banana singing the blues!

A defining hallmark of imagery is that, unlike perception, it is not driven by sensory input from the outside world. We can form a visual image of a car with our eyes completely shut. Imagery can be driven by information coming from memory, an internal information source. Having said this, though, numerous studies have shown that there are many similarities between perceptual processes and imagery. In particular, both processes seem to draw on the same underlying neural machinery. Farah (1988) showed patients who had suffered damage to the visual cortex as the result of a stroke and who could no longer form visual images. Goldenberg et al. (1988) showed that when participants are asked to answer questions that require the use of visual imagery, there is a greater increase in vision-related brain activity, measured as increased blood flow to the visual cortex, than in participants who are given questions that do not require the use of imagery.

It is important to note that "imagery" in a more general sense refers to the mental representation of something and need not be restricted to the visual domain. It is possible to have auditory, or olfactory, images as well. For example, one can imagine the melody to the song "Yankee Doodle Dandy" or imagine the smell of a steak being grilled over an open fire. As was the case with perception, most of the research that has been conducted in this area as well as theorizing on this subject have focused on the visual, so we'll discuss only visual imagery here.

The Kosslyn and Schwartz Theory of Visual Imagery

In this section, we devote much of our attention to an early and influential theory of imagery (Kosslyn & Schwartz, 1977; Kosslyn, 1980). It is a functional theory, meaning that it describes what the brain does with images. It is also a general theory; it attempts to account for the diverse operations that are involved in image processing, as opposed to its being a specific theory that focuses only on one type of operation. Mental rotation, described in this section, would be an example of a more specific theory. The Kosslyn and Schwartz theory of visual imagery was elaborated in enough detail that it has been implemented as a computer program. In describing this theory, Kosslyn

differentiates between the structures that are involved and the processes that act on them. In keeping with this, we will describe the structures first and the processes second.

Image Structures

Kosslyn posits two kinds of image structures. A **surface representation** is a quasi-pictorial representation that occurs in a spatial medium. It depicts an object or scene and underlies our experience of imagery. But what is this medium? He describes the medium as a **visual buffer.** This buffer is a surface matrix that consists of an array of points, similar to what makes up an image on a TV screen. The visual buffer functions as coordinate space and has a limited extent and a specific shape. It also has a grain, or limited resolution (Kosslyn, 1975). The resolution is highest at the center and decreases toward the periphery (Finke & Kosslyn, 1980).

Representations in the buffer are short-lived. Much like an image on a computer or TV screen, a visual buffer representation fades away if it is not refreshed. If an object with many parts is imagined, the parts generated first may decay before those generated later. This can explain why images of objects that are part of complex scenes are more degraded than images of objects in simple contexts (Kosslyn, 1975).

According to the Kosslyn theory, certain characteristics of the surface image are unique to analog representations. The images exhibit features not present in propositional representations. To start, the size of the image, shape of the image, its orientation, and the location of information in the image do not exist independently. When the value of one is specified, values of the others must also be specified. If we were imagining a fish in different ways, for example, as a big or small fish, a fat or thin fish, a right-side-up or an upside-down fish, or a fish to the left or right, then no part of the fish in the image is guaranteed to stay the same after the transformation. Images in this sense are holistic; a change in the image produces changes across its entire extent. In addition, any part or piece of an image is a representation of a part or piece of the actual object being represented, the referent. The part of an image of a fish that corresponds to the fin would represent that fin on the real fish.

What about the content of such images? Images are about objects. They are viewer-centered and represented according to the perspective of the observer who is viewing them. An image can represent an object directly, as a photograph does, or indirectly, as a diagram or schematic does. Content is determined not just by the image itself, but also by how the image is interpreted. The meaning of an image therefore does not lie entirely within the image itself, but also in how that image is processed by an interpretive device.

Deep representations constitute the second type of image structure. They consist of information in long-term memory that is used to generate the surface representations. Kosslyn describes two classes. **Literal encodings** contain lists of coordinates that detail the placement of points in the surface matrix such that the represented object is depicted. Some evidence suggests they originate in the right hemisphere. **Propositional encodings** are abstract, language-like representations, similar to declarative statements. They contain information about an object's parts, the locations of these parts, and their sizes. They are believed to lie in the left hemisphere. It helps to think of these encodings in the following way: Imagine that you receive a chair that requires assembly from an office supply store. The chair would come with a list of parts, such as the base, legs, and back, as well as the various nuts and bolts needed to fasten the parts together. This list of parts corresponds to the propositional encodings. The chair would also come with a set of instructions specifying how to connect the parts. It would tell you which bolt to use to fasten part A to part B. This instruction list is analogous to the literal encodings.

Image Processes

We have already reviewed the empirical evidence that supports each of the three types of image processes. **Image generation** occurs when the encodings in long-term memory are used to form an image in the visual buffer. Kosslyn envisions this process as a set of distinct operations, each labeled with a distinct command in the computer simulation. A PICTURE command converts the information that has been received from the encodings, FIND locates the position for the placement of a new part, PUT adjusts the size of the imagined part, while IMAGE coordinates other commands, and performs such operations as determining the appropriate level of detail.

One question concerning image generation is whether images are formed all at once or bit by bit. The evidence suggests that they are formed bit by bit (Kosslyn, 1980). Images that have more parts take longer to create. The same is true for images that have more detail. Data from another study show that people who are asked to create images have some control over how complete and detailed their images will be; depending on task demands and individual preferences, images can be sketchy or elaborate (Reisberg et al., 1986). This result is in accord with the idea that images are formed gradually rather than all at once.

Image inspection occurs when we are asked some question about an image, for example: "Is a pig's tail lower than its snout?" Inspection is like looking at the image with the proverbial "mind's eye." It also consists of a number of distinct processes, each labeled with a different command in the computer simulation. RESOLUTION determines if the image is at the right scale and

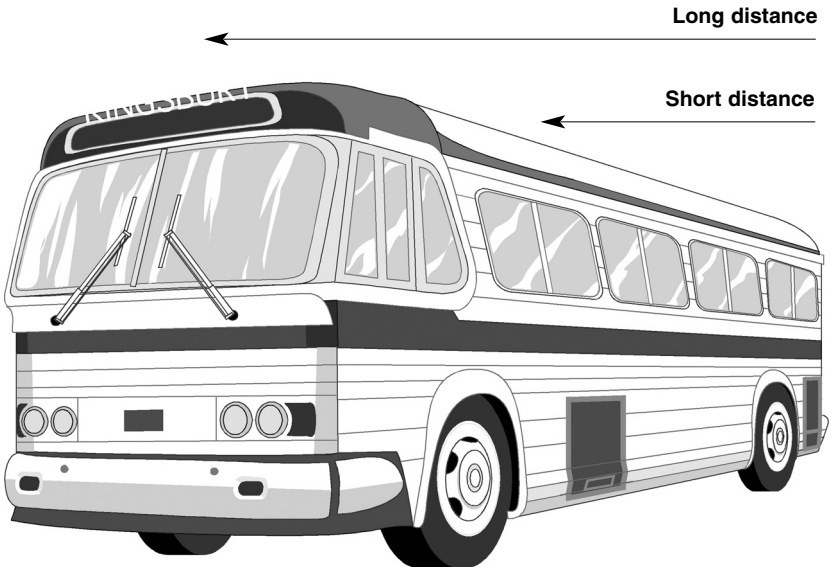


Figure 5.6 Study this picture of a bus for one minute. Then close your eyes and form a visual image of it, focusing on the back of the bus. Answer these two questions: Is the luggage compartment open? How many headlights are there? Which question took longer to answer? Why?

calls on ZOOM or PAN to expand or contract the image. SCAN, as its name suggests, scans between locations within the image.

An experiment by Kosslyn (1975) suggests we can zoom in on or magnify images as part of the inspection process. He asked participants to imagine a large object next to a small one. For instance, subjects were asked to imagine an elephant standing next to a rabbit. They were then asked to respond to questions, for example: "Does the rabbit have a tail?" Kosslyn found that, under these conditions, participants took longer to respond to questions about the rabbit. His explanation was that the rabbit was relatively small in the image and therefore difficult to resolve. Kosslyn inferred that participants zoomed in on or magnified the part of the image in question in order to respond.

Another study, by Kosslyn, Ball, and Reiser (1978), showed that images may be scanned or "looked across" in the same way real objects are. To get a sense of what participants were asked to do in this study, try the exercise that is depicted in Figure 5.6. The participants in the study were asked to memorize a map of a fictitious island. They were then instructed to form an image of the entire island and to focus on one of the objects on it. The name of a second object was announced, and participants were asked to scan their images by

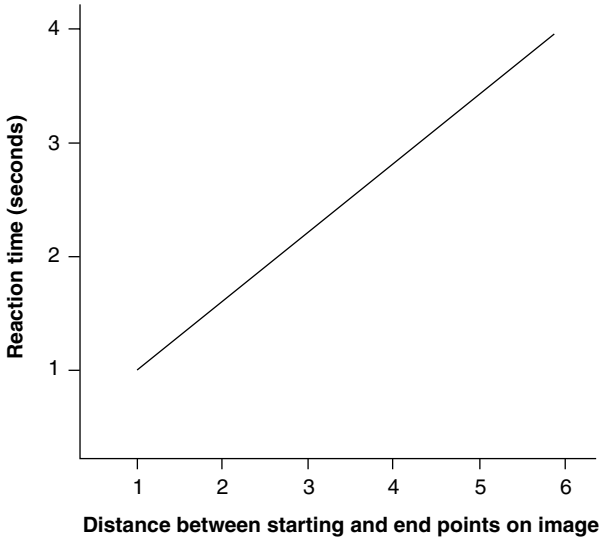


Figure 5.7 Idealized results from a visual image study showing time needed to mentally scan from one point of an image to another as a function of the distance between objects

imagining a black speck moving in a straight line from the first object to the second. When the black speck arrived at the second object, participants pushed a button, stopping a timer. When reaction times were plotted as a function of map distance, a linear relationship was revealed. Participants took longer to respond when the distances between objects on the island were greater. From this result, it was inferred that images do preserve spatial extent. Figure 5.7 shows the pattern of the results that have been obtained in experiments like this one.

Image transformation refers to an operation performed on an image. Kosslyn describes two transformation modes—shift transformations and blink transformations. Most image transformations are shift transformations, in which the image is altered incrementally. These changes can be seen as continuous changes in which the image passes through intermediate stages along a trajectory of size, orientation, or location. Blink transformations, in contrast, alter the entire image at once and can be used to re-form a complete image after it has decayed in the buffer.

A classic experiment in cognitive psychology calls attention to an example of a shift transformation. Shepard and Metzler (1971) presented participants with drawings of pairs of three-dimensional objects (see Figure 5.8). Sometimes

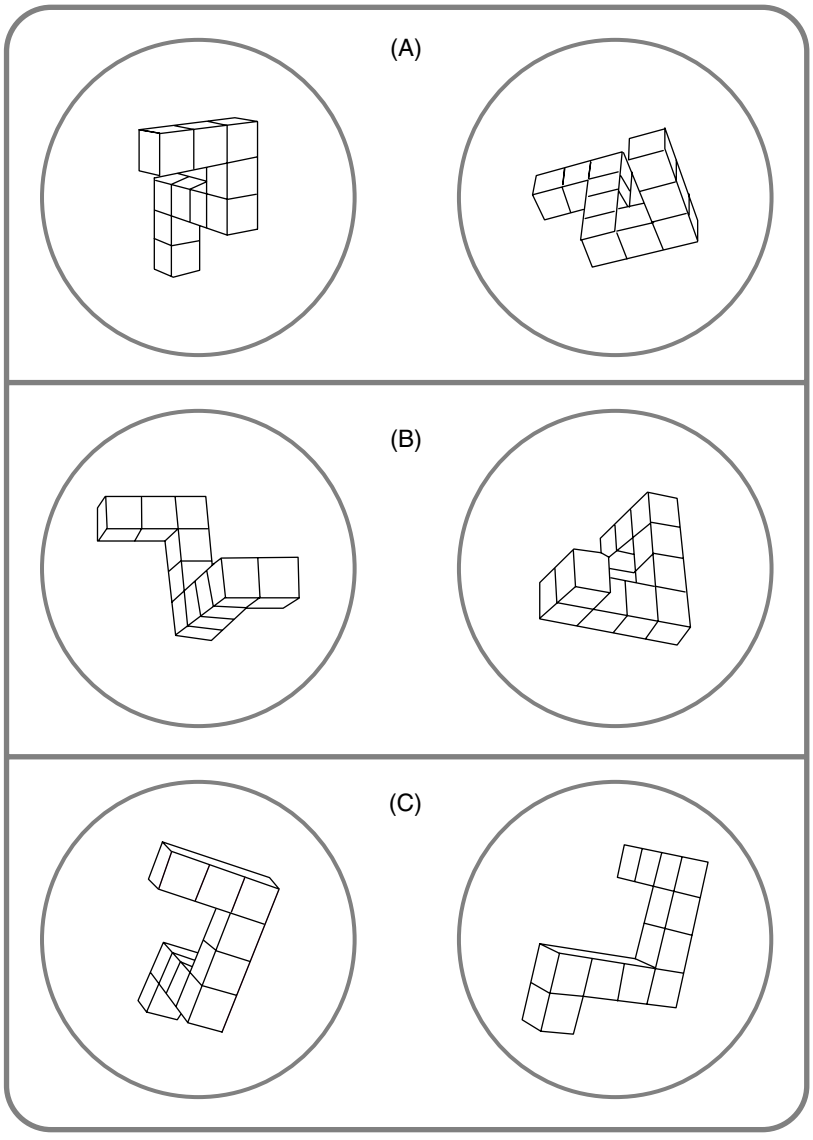


Figure 5.8 Stimuli from an early mental rotation experiment. Participants were asked to judge whether the two patterns were rotated versions of each other

Source: From Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 153*, 652–654. Copyright © 1971 AAAS. Reprinted with permission.

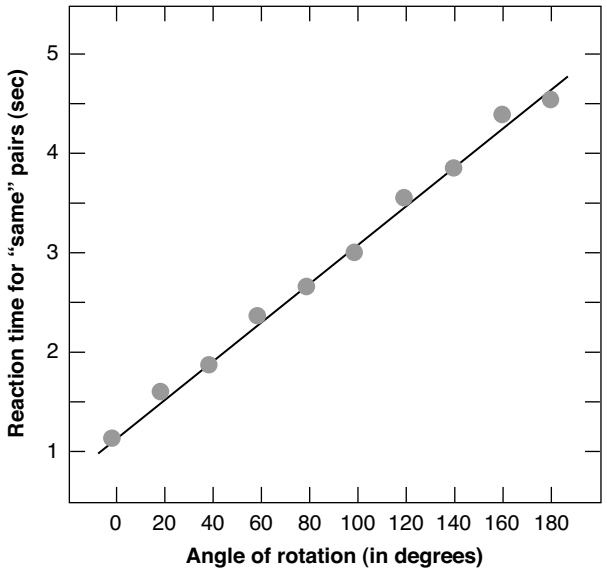


Figure 5.9 Response time plotted as a function of the angular difference between the two patterns suggests a mental rotation

Source: From Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 153,* 652–654. Copyright © 1971 AAAS. Reprinted with permission.

the paired objects were different. But sometimes they were the same, in which case the two objects were rotated versions of each other. The observers were asked to evaluate the pairs and to respond by pulling one lever, with their right hand, if the objects were the same, and another lever, with their left hand, if they were different. The response times were then plotted as a function of the angle of rotation (see Figure 5.9). The plot demonstrates that the greater the angular difference between the shapes, the longer it took to respond. Shepard and Metzler concluded that participants had formed a visual image of an object that was based on one of the objects in the drawings and then mentally rotated this image until it matched the other object.

Evaluating the Kosslyn and Schwartz Theory

Pylyshyn (1981) provides several critiques of the theory. He says there are too many free parameters, that is, functions or assumptions that are uncontrolled and that can affect the results of any experiment designed to test the theory. For instance, Pylyshyn points out that the transformation processes that are hypothesized to act on an image are very flexible. This flexibility,

rather than being a strength, leaves many questions unanswered, for example: Under what circumstances will a given transformation process operate? Pylyshyn would like to see a more detailed theory and a greater number of constraints put on its operation. He sees the theory as too open and as allowing for too many uncontrolled possibilities.

A second critique concerns **cognitive penetration**, in which one's knowledge, beliefs, goals, or other cognitive states alter performance of a mental task. Pylyshyn argues that mental imagery processing is biased by people's expectations. In Pylyshyn's view, our knowledge of what objects are like and the way they act in the world influences the way we process images of these objects. For example, if we think an object is heavy, we may mentally rotate it more slowly. If this is the case, and if imagery operations are cognitively penetrable, then they cannot be, as suggested by the theory, innate, universal, or automatic.

Pylyshyn also makes the argument that participants performing a mental rotation task may not be rotating at all. They may in fact be employing non-analog or propositional means of carrying out image transformation. The data that appear to support image rotation are not reliable, according to Pylyshyn, and have been obtained because the task and the instructions led participants to recreate, as accurately as they could, the perceptual events that would have occurred if they had been observing an analog process. In other words, the observers are doing what the stimuli, task, and experimental situation seem to demand of them. This phenomenon is believed to be a potential source of error in many psychology experiments and goes by the name of **demand characteristics**.

Kosslyn counters these three arguments. He makes the distinction between parameters and free parameters, and he points out that his theory contains many of the former, but not many of the latter. This theory, or any general theory that explains new cases, must include many parameters. Specific theories have fewer parameters, but can explain only a limited number of cases. With regard to cognitive penetrability, Kosslyn admits that rotation and other image transformation processes ought to be cognitively penetrable, otherwise the processes could not be applied to solve problems or could not be used in novel ways. Finally, Kosslyn points out several experiments in which task demands affected the results, but argues that these studies in no way preclude the use of analog imagery processing.

The Imagery Debate

In the 1970s and 1980s, not long after the outbreak of the cognitive revolution, there was considerable debate in academic circles as to whether or not

visual images even exist. As Pylyshyn suggests above, many of the tasks subsumed under the heading of imagery could in fact also be performed by propositional operations. Ned Block (1981) sheds some light on the two sides in this controversy. According to Block, people who believe in visual images are pictorialists. They don't believe that we literally have pictures in our heads, but they do believe that we form quasi-pictorial representations that represent in the same way that pictures do. Pictorialists advocate the widespread use of mental analog representations. On the other side there are descriptionalists, who believe images are not pictorial. The subjective experience of having a visual image makes it seem as if there is a picture "in there," when in reality it is all just digital symbol processing.

This debate has some important implications. As we mentioned in the introductory chapter, analog and digital representations lend themselves to different types of processing. Images are processed spatially; they are subject to transformative processes such as zooming, scanning, and rotation. Digital symbols are processed according to syntactical rules similar to those of grammar or arithmetic. Because traditional computers employ symbols and processing of the digital sort, the existence of mental images would imply that at least some human mental operations differ fundamentally from computer operations. This means that in order to truly understand these mental phenomena, we would probably have to design computers with analog computational architectures. These computers would then give us new insights into mental processing that are different from the ones provided by simulating analog processes on a digital computer—something that Kosslyn and others have done.

We have already presented much of the evidence in the case for imagery. This includes studies that show the similarity between imagery and perception and their shared neural machinery. It also includes the many experiments that provide evidence for various forms of analog processing. So let us look at some more arguments against imagery.

Daniel Dennett (1981) stands on record as a descriptionalist. He states that the only place where we seem to have an image inside the head is the retina, a place where a two-dimensional mapping between stimulus structure and neural activation occurs. From that point on, according to Dennett, there is no neural "map" that seems even remotely like an image. A representation of a visual scene in the occipital lobe is a very distorted representation of the original stimulus. The image has been split into left and right visual field halves that project onto both hemispheres, and, within each hemisphere, onto brain areas that represent one eye or the other. The pictorialist's response to this is that the image need not be a perfect pictorial representation (hence the pictorialist's use of the term *quasi-pictorial)*. All that is needed is a representation that in some

way mimics the spatial properties of the stimulus. Representations of this kind can be instantiated in the brain in many ways.

Dennett points out that mental images are vague, yet real images are not. He gives the example of a tiger. A picture of a tiger has an exact number of stripes, whereas an image of a tiger does not seem to. The "vagueness" of visual images implies that they are propositionally coded, as language is better suited to expressing ambiguity such as is expressed in the phrase "a few stripes." Pictorialists counter this by saying that images can have any degree of clarity. An image, if viewed from a distance, will have less detail and will appear fuzzy. But a zooming operation can increase the resolution, allowing us to count stripes if necessary.

The current emphasis in this debate is not so much on whether images and image processing exist, but on which types of tasks call on them and under what circumstances. It is now generally acknowledged that analog *and* digital processing, as well as pictorial *and* descriptional processing, both happen. According to dual-coding theory, discussed in the introductory chapter, items can be represented by both verbal and visual codes (Paivio, 1971). The ability to remember concrete items such as "cat" is better because we have both a verbal code and an image to represent such an item. Memory for abstract words such as "justice" is poorer because it is only possible to form a verbal code for an item of this kind.

Problem Solving

Problem solving is the transformation of a given situation into a desired situation or goal (Hayes, 1989). Problems are solved not by people only, but also by computers. The former type of problem solving is within the purview of the cognitive approach and is studied by experimental psychologists. Researchers in artificial intelligence and robotics study machine problem solving. We will postpone a discussion of their perspective on problem solving until later in the book.

We have already discussed two other approaches to problem solving. The first approach was described in Chapter 1 and centers on the use of analogies. There we presented evidence that showed that the structure of one problem could be applied to and used to solve another. The second approach was the Gestalt perspective, described in Chapter 3. You may recall that the Gestalt psychologists focused on insight learning, in which a solution suddenly manifests itself, perhaps as a result of some amount of subconscious processing. In this section, we focus on an approach to problem solving that has been

influenced by the discipline of artificial intelligence. The approach relies on the setting of subgoals and the testing of different possible solutions in the pursuit of a solution.

Anderson (1980, 1985) lists four characteristics of problem solving:

1. *Goal directedness*. Problem solving behavior is directed toward the attainment of a **goal**, which is the desired end point or solution to the problem. If we were attempting to solve an arithmetic problem, say, 25 + 36, our goal would be to find the sum of these two numbers.
2. *Sequence of operations*. This problem solving behavior involves a sequence of steps. To add 25 + 36, we can first add 20 + 30, to obtain 50. That would be step one. We would then add 5 + 6, to get 11, in step two. We would then add 50 + 11, to get 61, this being the third and last step.
3. *Cognitive operations*. This is the application of some process to the problem solving situation that transforms the situation. Each permissible cognitive action that is applied in this way is called an **operator**. Operators in arithmetic include the addition of two numbers, the carrying of a remainder, or the retrieval of a number from memory.
4. *The setting of subgoals*. Each step in the problem-solving sequence produces a **subgoal**, an intermediate goal that is set along the way to eventual solution of the problem. There is thus a hierarchical organization of goals in problem solving. The overall goal is broken down into subgoals. Each subgoal may in turn be broken down further, into additional subgoals. In our example, adding 20 + 30 yielded 50, in step one. This would generate the subgoal of adding 5 + 6, which produces the new subgoal of adding 50 + 11 in the pursuit of the final goal, the total sum.

The problem space is a very important concept in problem solving. A **problem space** may be thought of as the different situations or states that can exist in a problem. For any problem space, one can define three important states. The initial state is what the problem is like at the outset. Intermediate states are what results after some action is applied to the problem. The goal state is a desired situation and can include the final situation or configuration of the problem after a solution has been obtained. A problem space also includes the problem solver's knowledge at each step of the problem solving sequence, knowledge that is applied at the step being negotiated, as well as knowledge that is in memory and could be applied. Problem spaces are usually represented in a diagram, called a solution tree, that shows all of the possible steps that can be taken in pursuit of a solution (see Figure 5.10). The solution tree contains

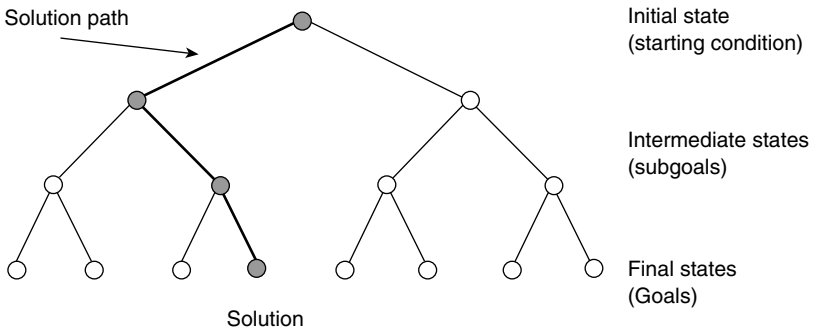


Figure 5.10 A representation of a problem space showing the solution path

a series of branchings. Each fork that has two or more branches represents the possibility of choosing one course of action over another. A given path can ultimately dead-end in a nonsolution. The correct solution to a problem is a particular path, or a set of paths if there is more than one procedure that produces a correct solution.

One approach to problem solving is to try all possible paths represented in the solution tree. This brute force method is computationally intensive and time-consuming. As an example, the game of chess has been estimated to have 10^{20} states, or possible moves a player can make. It is simply not possible to search through all these states to locate the right one. Current computers can search several billion states, still far fewer than the total number of states. Chess grandmasters usually don't consider more than 100 and an average player might only consider several.

What is needed then is some guide to tell us which part of the space is likely to yield a correct answer. Heuristics serve this function. A **heuristic** is an informal "rule of thumb" method of problem solving that does not guarantee a solution but is faster and easier to use than a systematic search. Imagine that you have misplaced your keys somewhere in your apartment. You could engage in a systematic search and look everywhere for the keys: in each room, under the bed, on the shelves, behind the drawers, and so on. Alternatively, you could look first in those places where you usually put the keys down: in a desk drawer, on a tabletop near the door, or some other likely location. This heuristic would usually enable you to locate the keys more quickly and with less effort.

The most effectual heuristic in problem solving is the **means-end analysis**. In this technique the problem is solved via successive determinations of the difference between the existing state and the goal or subgoal state, and then the

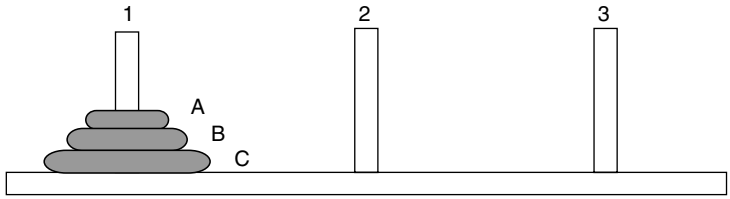


Figure 5.11 The three-disk Tower of Hanoi problem

identification of and utilization of an operator that reduces this difference. We all use means-end problem solving in our everyday life. If you wanted to get groceries for the week (the goal), you would need to first get to the supermarket (a subgoal). Since you are at home, you would need to drive to the supermarket (an operator). If this were accomplished, the first subgoal has therefore been attained, and the difference between the state of being at home and that of needing to get to the market is reduced. Once you are at the market, you would set up another subgoal of getting the food, which would require a second operator, obtaining a shopping cart. This process would then continue until the final goal has been reached.

The Tower of Hanoi problem has been studied quite thoroughly by investigators and serves as a good example of the application of the means-end analysis. In this problem, there are three pegs, marked, from left to right, 1, 2, and 3 (see Figure 5.11). Initially, there are three disks of different sizes, A, B, and C, that have been stacked on peg 1—the largest on the bottom, the smallest on top. The goal is to get all three disks on peg 3, stacked in exactly the same way. There are several constraints: you can move only one disk at a time and only to another peg. You can also only place a smaller disk on top of a larger one. Try to solve the Tower of Hanoi problem before reading on.

Figure 5.12 shows the seven-step solution. The first goal is to free up disk C, since it must always be on the bottom. Doing this entails setting the subgoal of getting disk B off of disk C. However, this necessitates the setting of another subgoal, getting A off of B. The first move is thus to transfer A to peg 3, which then allows us to move B to peg 2. The next few moves are along these lines, with the generation of new subgoals and new operators for their attainment. We cannot move C to peg 3 just yet, because A occupies that position. To remove A, we place it on top of B on peg 2. This frees up peg 3, so that we can now place C there. The remaining moves are straightforward. A goes to peg 1, B goes to peg 3, on top of C, and A can finally be placed atop B to complete the attainment of the goal.

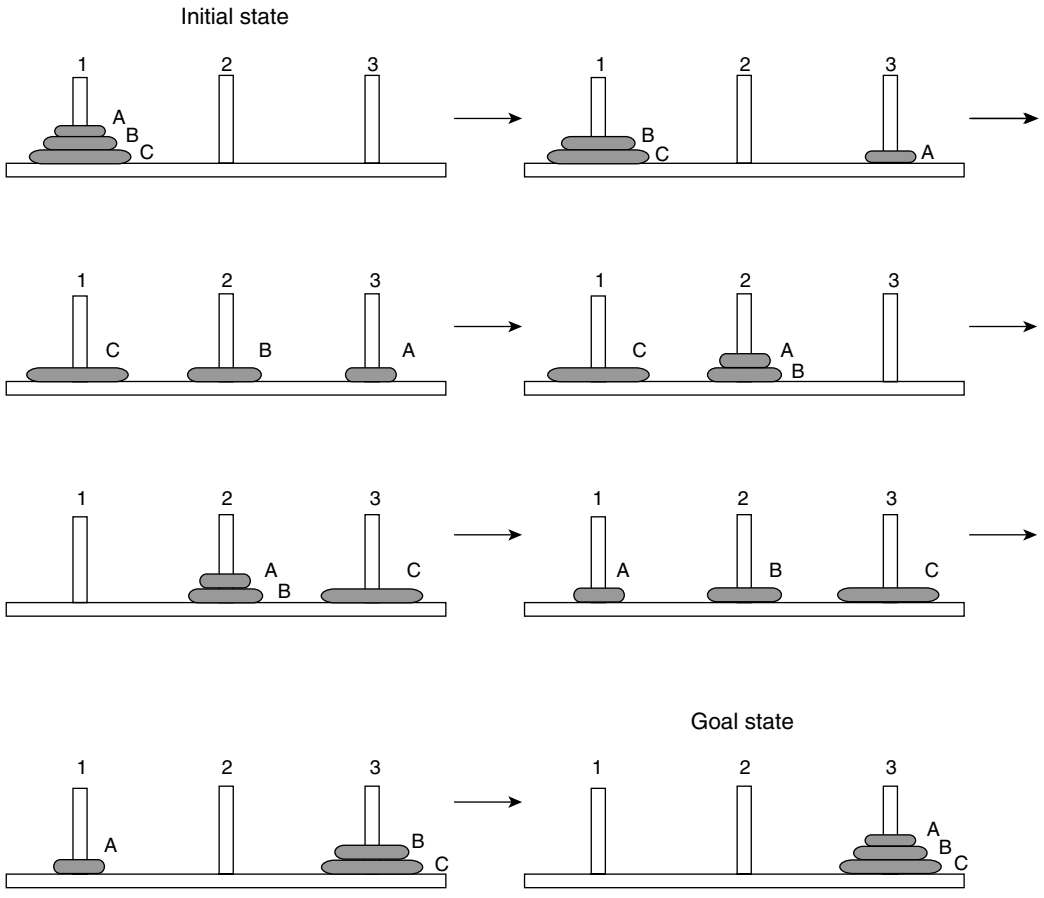


Figure 5.12 The seven-step solution to the three-disk Tower of Hanoi problem

The General Problem Solver Model

Alan Newell and Herb Simon were early pioneers in problem solving research. They developed the first comprehensive computer simulation of human problem solving (Newell & Simon, 1972). Their program applied means-end analysis to a wide range of problems and thus the program earned the name General Problem Solver (GPS). The GPS model serves as a useful analogy for how humans sometimes solve problems—by first representing the problem and the goal, then identifying the subgoals and operators that will produce that goal. The model suggests that humans and computers can both

solve problems this way—by constructing "plans within plans" to effect a solution.

A key component of the model was its use of production rules. These rules were used to perform the operations that were required to bring about a solution. Ashcraft (2002) gives some examples of production rules for the Tower of Hanoi problem:

1. IF the destination peg is clear and the largest disk is free, THEN move the largest disk to the destination peg.
2. IF the largest disk is not free, THEN set up a subgoal that will free it.
3. IF the subgoal of freeing the largest disk has been set up and a smaller disk is on that disk, THEN move the smaller disk to the stack peg.

The GPS program has been applied to several well-known logic puzzles, including the Tower of Hanoi problem, the Missionary-Cannibal problem (described later in this chapter), the three coins puzzle, the water jug task, and more. It was able to solve each of these. In addition, Newell, Simon, and others have tested the model. In one study, they had participants talk out loud as they attempted to solve abstract logic problems (Newell & Simon, 1972). They found that there was a significant correspondence between the steps the participants reported using in their verbal protocols and the steps employed by GPS.

Evaluating the General Problem Solver Model

However, GPS has its limitations. In some cases, it diverges from the strategies human beings use to solve problems. With some problem solving, human beings are apt to use another heuristic, called the hill-climbing strategy. It involves taking actions that always bring one closer to the goal. To illustrate, if you were climbing a mountain and your goal was to get to the top, you would always choose a path that led upward. This strategy can sometimes fail, however—in this example, in the case of a path that goes downward and then back upward toward the summit. In the Missionary-Cannibal problem, there is a necessary step where one must carry people who have been transported to the far side of a river back to the original side of the river. Because this leads away from the final goal, which is to move everyone to the far side, persons attempting to solve the problem do not take the step, and therefore fail to solve the problem. The GPS program doesn't have difficulty with this particular kind of pitfall, but does sometimes fail to obtain a solution because it applies the means-end heuristic too rigidly (Greeno, 1974).

The SOAR Model

SOAR, which historically stood for State, Operator And Result, is a problem-solving model, but it is also a universal cognitive architecture, a system designed to account for a wide range of cognitive phenomena (Newell, 1991). **Cognitive architectures** specify the structure and function of many different cognitive systems and how the structure and function interact. Broadbent's (1958) early model of attention, the Atkinson and Shiffrin (1971) modal model of memory, and Anderson's (1983) ACT* memory model are all examples of other cognitive architectures. These models can be applied within a specific domain, such as attention, memory, or problem solving, but they go beyond this, as they attempt to describe basic principles of cognitive functioning.

As we will see SOAR incorporates some of the elements found in GPS, but is more general. Let's discuss its main features. SOAR represents all its tasks as problem spaces. Any problem is broken down into an initial state (the problem's initial starting point), a desired state (the goal or solution), and the current state (what the problem is like at any given moment). It can apply any one of a set of operators to the current state, thereby altering it to create a new state that is closer to that of the desired goal. SOAR was tested on block-world problems. A block-world problem consists of a set of blocks that has one configuration, a starting configuration, which must be rearranged so as to yield a desired end-state configuration (see Figure 5.13).

In SOAR, knowledge stored in a long-term memory is used to select operators and guide movement through the problem space. The memory consists entirely of production rules that specify an action that is to be taken if a set of preconditions is met. Objects, such as blocks in the block-world problem, or disks in the Tower of Hanoi problem, are represented by attributes and values. These are simple descriptive characteristics that can be used to define objects in the problem scenario exclusively.

SOAR goes through a decision cycle, where it accumulates evidence that has to do with the problem. Once this has taken place, a decision is executed. Examples of decisions that can be made are which problem space is to be used, which state is to be used within a given space, and which operator is needed to enable progression to the next state. One of the model's most interesting aspects is its capacity to develop preferences as it is accumulating evidence. These preferences are for actions that ought to be taken, given their likelihood of bringing about the goal. Based on these preferences a given operator may be accepted, rejected, or considered as better, indifferent, or worse.

SOAR, like other computer programs, can follow a preprogrammed set of instructions for solving a problem. But, unlike these programs, it is also capable of generating new ways of approaching the problem. SOAR creates novel

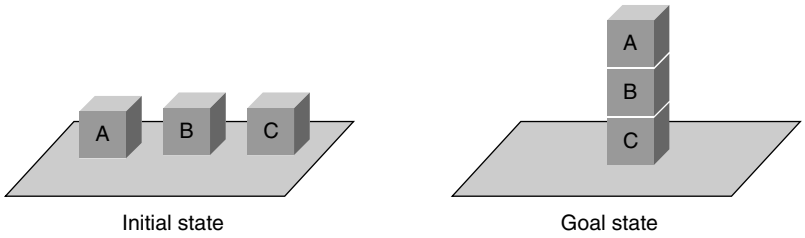


Figure 5.13 A typical block-world problem

subgoals when it comes up against an impasse—a situation in which there is no option better than the rest. The setting of the new subgoal is based on knowledge stored in memory. This ability allows SOAR to adapt itself to the demands of a novel situation flexibly and dynamically.

Figure 5.14 gives an example of how SOAR would solve a simple block-world world problem. The initial state has blocks A, B, and C arranged in a row on a table. The goal state is a reconfiguration of the blocks such that block A is on top of block B, and block B is on top of block C. At the initial starting point there are three permissible operators. A can go on top of C (A → C), B can go on top of C (B → C), or A can go on top of B (A → B). SOAR evaluates these three options by trying them out in an evaluation space. The A → C option produces a configuration that leads to an unacceptable solution, with B on top of A and A on top of C. The A → B option also produces an unacceptable state, with C on top of A and A on top of B. SOAR then selects and implements the only remaining operator to produce a successful solution.

A final comment on the SOAR model concerns chunking, which in this context is comparable to the process of grouping items together in memory, discussed earlier. Chunking in this context corresponds to learning from experience. If SOAR has tested possible solutions and found a particular solution to a problem in the past, it stores this solution in a general form in its memory. When SOAR encounters a similar problem at a later date, it can then apply operators it has applied previously to effect a similar solution. SOAR in effect recognizes the problem as a combination of conditions that trigger the appropriate action. This grouping or chunking of the conditions into a single precondition for applying an operator gives the process its name.

Evaluating the SOAR Model

The SOAR program represents an ambitious attempt to develop a cognitive architecture that does more than solve problems. It was designed as a

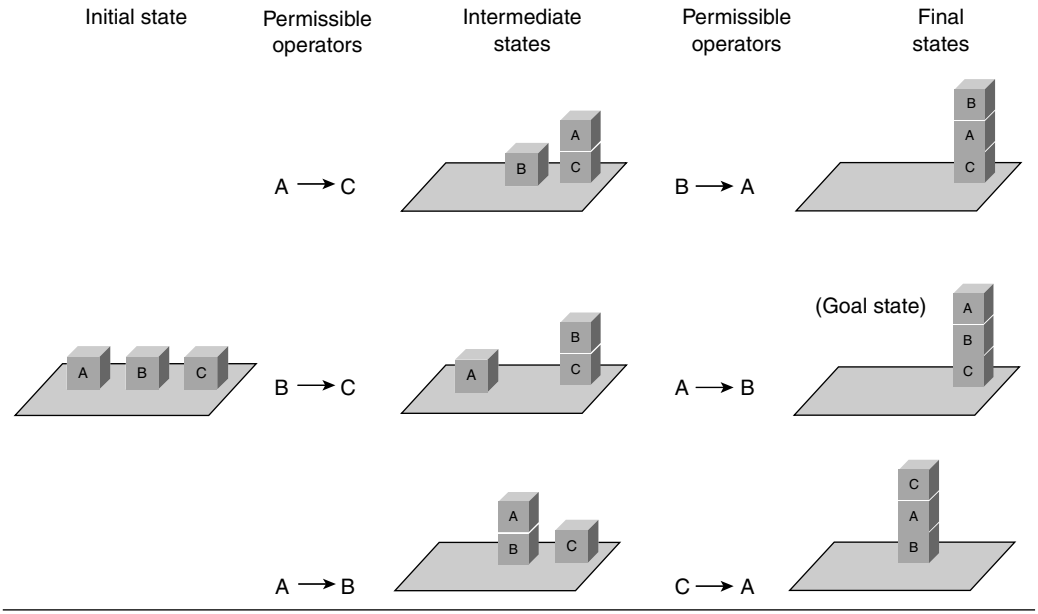


Figure 5.14 How SOAR might solve a block-world problem

model for delving into the general principles that underlie learning, memory, problem solving, and the interactions among these systems. It has a good track record, having been applied to and having solved a variety of problems.

What's more, the incorporation of the chunking feature parallels the way human experts diagnose and solve problems. Chase and Simon (1973) presented pictures of a chess board whose pieces occupied random positions and pictures of a game in progress to two groups of participants—expert chess players and novice chess players. There was no difference between the recall of the novice players and that of the expert players when it came to remembering the pictures of the randomly placed chess pieces, but experts were better than novices when it came to remembering the game-in-progress pictures. This suggests that the experts were chunking information; they were organizing portions of the information into meaningful wholes. These wholes then guided the selection of the appropriate operator to be applied to the problem. For the expert chess players, seeing a particular configuration of pieces apparently triggered a memory of how that situation was dealt with in the past. Auto mechanics and physicians also chunk when they diagnose the cause of a mechanical breakdown or biological illness from a set of symptoms.

A critique of the model centers on the assumption that only production rules make up memory. Other architectures, notably ACT*, posit separate memory types with separate structures that code for declarative and procedural knowledge. The SOAR view of a single unitary type of memory simplifies the model, but runs contrary to evidence that supports the existence of distinct memory systems.

Overall Evaluation of the Cognitive Approach

The strengths of the cognitive approach are many. Unlike the disciplines that preceded it, cognitive psychology is governed by a set of common assumptions. These are that the mind represents and operates on information and that specific models of how the mind does these things can be formulated. Another advantage of the cognitive approach is that, rather than view the mind in a general, undifferentiated way, as some of the adherents of other approaches did, it compartmentalizes the mind into distinct processes. This carving up of the mind into modules makes research easier because it simplifies and delimits the range of a phenomenon under investigation. Although not all cognitive processes may actually be compartmentalized, this compartmentalization has usually proven to be a useful approach.

Another advantage of the cognitive approach is, as was discussed at the end of the previous chapter, its use of model building. Models provide a concise way of formulating how a given mental process might operate. They are more precise than theories and can be tested using experimental data. The implementation of a theory as a model and the subsequent modification of the model on the basis of empirical findings is a fruitful method for uncovering more about a cognitive process.

Of course, like any other discipline, the cognitive approach has its limitations. It supports the traditional view of information processing, in which representations are symbolic and transformations are operations that are performed on the symbols. This needs to be reconciled with the network view—that information is coded in a distributed way and that transformations consist of activations of nodes in a network. Model builders in cognitive psychology also need to consider findings from neuroscience. Any model maker that attempts to provide an information processing explanation of a brain function should map the different components of the relevant model onto the known anatomical and physiological evidence. This will help to delineate and develop the model so that it more accurately reflects its biological implementation.

In Depth: Search in Working Memory

You go to the grocery store to buy some food for dinner. Because you needed only a few items, you memorized the items rather than wrote out a list. After buying rice and broccoli, you pause to consider what else it is you have to purchase. What mental process occurs when we perform this sort of task? The process in question is called "search," and it is one of the many operations carried out in working memory. In this section we describe a classic experiment from cognitive psychology that suggests how such a mechanism might work.

Search, in an abstract informational sense, involves looking for a target among a set of items. Each of the items must be represented mentally, as must the target. Then, a representation of the target must be compared against each of the represented items in the set in the expectation that a match will be made. But now a question arises. This search process can proceed in two ways. If it is a **serial memory search**, then comparisons take place one at a time, with a participant perhaps beginning at the beginning of the list and proceeding through till the end. If this is the case, then search times should be long, as each comparison takes time. If it is a **parallel memory search**, then the target can be compared to all the items at once. Assuming this to be the case, search times should be short, as only one comparison is being made.

To find out whether search was serial or parallel, Sternberg (1969) asked participants in his study to memorize a set of items, such as a series of digits. The size of the set varied between one and six digits—always below the 7 ± 2 limit. Sternberg then presented a target digit. The participants had to decide whether the target was a member of the memory set or not. For example, in one trial that used a set size of four, the participants might have had to memorize the digits 4, 9, 0, and 2. They were then given a target, the number 9. Following presentation of the target they were to respond as quickly as possible, pushing one button if the target was present in the set (a positive trial) and another if it was absent (a negative trial).

Reaction time (the time it took to push the buttons) was analyzed and plotted against set size for both positive and negative trials. Sternberg made two predictions. If the search was serial, he should obtain an increasing function with longer search times for larger set sizes. This follows, since if each comparison takes a given amount of time, a greater number of comparisons will require more time. If, on the other hand, the search was parallel, he should obtain a flat function over set size, indicating the smaller amount of time it takes to compare the target to all items in the set at once.

Sternberg made yet another prediction. He speculated that search, if it were serial, could be self-terminating or exhaustive. In a **self-terminating search**, the

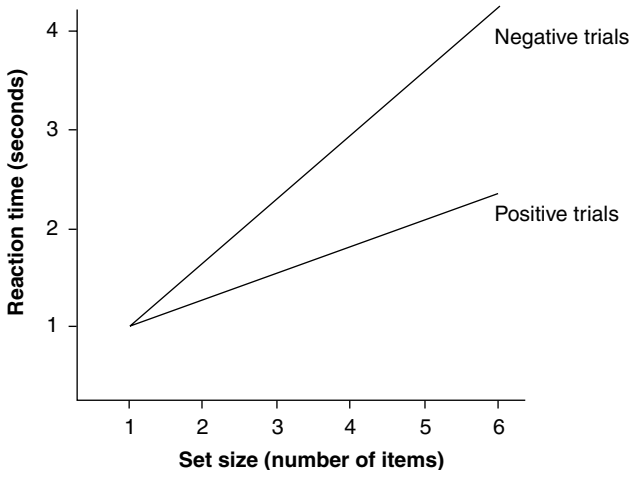


Figure 5.15 Idealized results from a memory scanning experiment showing that scanning of items in working memory is serial and exhaustive

search process stops as soon as a positive match occurs. Because the target can be located anywhere in the set, the search times will be variable. Sometimes, if the target matches an item at the beginning of the list, the search will end quickly. On other occasions, when the target matches an item that is near the end of the list, it will take longer. This effect manifests itself as a lesser slope in the search function for positive trials, in which a target is present in the set, but steeper slopes for negative trials, in which scanning must always continue to the end. In an **exhaustive search**, comparisons are made all the way to the end of the list, even if the target has already been located. In this situation, the slopes for positive and negative trials should be both steep and identical.

The data showed that search was serial and exhaustive. The search function increased with set size and had a lowered slope for positive trials (see Figure 5.15). At first blush, this result seems counterintuitive and inefficient. Why should you continue looking for something you've already found? There are several possibilities. Exhaustive search could serve as a "double check," a way to ensure that the target isn't duplicated later in the list. Or, it could be that once search is initiated, it proceeds automatically and cannot be stopped. The latter possibility is interesting because it implies that search is only partially under voluntary control. The central executive in the working memory model may issue a command to the articulatory loop to begin a search, but once this command is issued it cannot be further regulated.

Minds On Exercise: Memory Effects

This exercise requires the making up of two lists of words. Each list will contain 20 words. Put only concrete words into the first list, words for which a visual image and a verbal representation are possible. Put only abstract words into the second list, words for which there is only a verbal representation. Make sure that all the words are fairly similar with respect to the number of syllables and the frequency of usage. Read off each list slowly to a group of participants. Then have them perform an immediate recall task; ask them to write down as many of the words as they can remember, in any order. When this has been done, read off the lists again and have the participants tally up the numbers of words they recalled successfully in each condition. Calculate the group's overall correct response rate, for both lists. For which list was the recall greater? Why? Ask the participants which strategies they used to remember in both cases.

If you have time, you can calculate average percent correct for the group as a function of the position of the words in a list. Was recall greater for words located at the beginning, middle, or end of the lists? Did this pattern differ in the concrete or abstract conditions? The **primacy effect** predicts that more words at the beginning of a list will be remembered because they have been rehearsed more often. The **recency effect** predicts greater recall for items near the end of a list because these items have not had time to decay from working memory. Words in the middle of a list are generally not remembered as well. One explanation is that they receive both proactive interference, from words near the beginning, and retroactive interference, from words near the end.

Food for Thought: Discussion Questions

1. In what ways might semantic and episodic information be organized? Do we store and remember facts in a manner different from the way we store and remember personal events? How might these two organizational systems differ?
2. How can you improve your memory? Compare your ideas to techniques for memory improvement that have already been established—called mnemonics.
3. When do you use visual imagery? When do you use verbal or propositional reasoning? Are there some tasks for which either could be used? Do you prefer one mode of processing to the other? Why?
4. Name processes other than rotation that can be applied to an image. Can the image inspection processes of scanning, zooming, and panning, and so on be considered transformations of an image, for example, a rotation? Why or why not?

5. Here is the Missionary-Cannibal problem. Three missionaries and three cannibals are on one side of a river and need to cross to the other side. The only means of crossing is a boat, and the boat can hold only two people at a time. Devise a set of moves that will transport all six people across the river, bearing in mind the following constraint: The number of cannibals can never exceed the number of missionaries in any location, for the obvious reason. Remember that, for every return trip, someone will have to row the boat back across the river. Hint: At one point in the solution, you will have to send more people back to the original side than you have just sent over to the destination side. Draw a diagram to help you solve this problem. Apply a means-end analysis to your solution. What is the goal? What are the subgoals and operators?

6. Bill is driving home from work at the end of the day. He runs over some broken glass and gets a flat tire. Draw the problem space for this problem. The drawing will include the solution path and the area of the space that would be reduced through the application of a heuristic.

Go to the website:
http://rec-puzzles.org/logic.html

Try your hand at a few of the puzzles. Attempt to solve the black hat "smullyan/black.hat" problem. How did you first try to solve it? Now, draw diagrams that depict the structure of the problem. Do these diagrams help you to determine a solution to the problem? Why or why not? If not, would a different diagram have helped you? How?

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Block, N. (Ed.) (1981). *Imagery*. Cambridge, MA: MIT Press.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, MA: MIT Press.

Schacter, D. L., & Tulving, E. (Eds.). (1994). *Memory systems*. Bradford Books.

6

The Neuroscience Approach: Mind as Brain

Referring to the brain: "My second favorite organ!"

—Woody Allen, 1973

The Neuroscience Perspective

Neuroscience is the study of nervous system anatomy and physiology. It is concerned with both the structure of this system in humans and other animals as well as its function. As such, neuroscience provides a body of knowledge that serves as a foundation for understanding how cognitive operations are carried out. It gives us a hardware or implementational level of description upon which we can base an algorithmic and computational description. Neuroscience is invaluable in this regard because it constrains the sorts of models of brain function that can be formulated in the other approaches.

There has been a fairly recent trend in neuroscience toward the integration of biology with cognition. Out of this union, a new discipline has emerged, called **cognitive neuroscience** or neuropsychology. The goal of this discipline is to explicate the structures and physiological processes that underlie specific cognitive functions. We adopt a cognitive neuroscience perspective in this chapter and focus on the cognitive operations that were surveyed in the two

chapters on the cognitive approach (Chapters 4 and 5). These cognitive operations are visual object recognition, attention, memory, and problem solving. This neuroscience perspective builds and expands on the information processing models for these operations that were described previously. For neuroscientific accounts of visual imagery and language please see the relevant sections in Chapters 5 and 9, respectively.

A fair amount of background knowledge is needed to interpret the findings of neuroscience studies. We will therefore provide general information on methodology, anatomy, physiology, and terminology prior to a discussion of the specific topic areas. Our present discussion begins with the tools used by neuroscientists, and is followed by a discussion of the geography and function of neurons and of the overall brain. We will then address the neuroscience of cognitive functions and models of their operation.

Methodology in Neuroscience

There are many procedures for collecting data in neuroscience. In this section we group these procedures into three main categories. The first category of procedures involves the investigation of brain damage. In studies of humans, researchers, using the case study method, investigate the behavioral deficits of patients who have suffered damage to some part of the brain as the result of accident. In animal lesion studies this damage is deliberately induced. The second category of procedures involves the recording of brain activity in healthy subjects. The electrical action of individual neurons or small groups of neurons can be measured with cellular recording techniques that use an implanted electrode. Larger patterns of brain activity can be measured through the use of surface electrodes that have been implanted on the scalp or through the use of more sophisticated brain-imaging devices. The third category involves direct stimulation of the brain itself.

Techniques for the Study of Brain Damage

It is of course unethical to damage a person's brain deliberately and then look at the effects. For this reason researchers examine brain damage and its effects in people that have come about as a result of accident—what is called the **case study** method. There are different types of brain damage. In a patient who has had a stroke, a blood vessel servicing a part of the brain bursts, depriving that part of the brain its blood supply. Direct mechanical damage to the brain accompanies head trauma. This type of damage can be incurred in an

automobile accident or a fall. Brain damage is also an end result of exposure to carbon monoxide gas. Usually in these cases, localized damage occurs and follow-up evaluations of affected persons can reveal exactly which areas were damaged. The resulting behavioral and cognitive deficits can then be mapped against the damaged areas. If, for example, the front part of the brain has suffered damage and the patient has subsequently had trouble planning, we could conclude that the front part of the brain is in command of the faculty of planning.

A related method of study involves the deliberate destruction of brain tissue or brain areas in animals and the examination of the resulting behavioral deficits. A brain lesion is a wound in or injury to brain tissue. A study that uses this method is thus referred to as a **lesion study**. Lesions can be generated in several ways. Brain tissue can be drawn off by suction, destroyed by passing a high-frequency radio current through an electrode, or frozen by pumping coolant through an implanted probe. With this last procedure, the tissue is preserved and normal function returns as the affected area warms. The motivation for using these procedures is the same as that for using the case study, and this type of study suffers from the same flaws.

Evaluating Techniques for the Study of Brain Damage

There is logic behind case studies and lesion studies. If brain area X is damaged and a deficit is subsequently observed in behavior Y, researchers infer that area X plays some role in the control of behavior Y. This seems straightforward enough, but is actually fraught with difficulties. Some have likened this kind of conjecture to removing part of a car engine to see how its removal affects the car's function. If we removed the carburetor, the engine wouldn't run because the spark plugs wouldn't receive any gasoline. However, the same would be true if we removed the gas tank. The point is that in a car many of the systems are interdependent. Removing any of a number of different parts can produce the same symptom. Areas of the brain are also interdependent. The effects of damage to one area could have a variety of functional interpretations. That area could be entirely responsible for a cognitive ability or part of a collection of areas that are in charge of that ability. Alternatively, the area could be a center where information is processed or simply a region through which pathways that connect one center to another pass.

A second criticism specifically of lesion studies is ethical instead of procedural. It has to do with whether animals should be used at all for research. This criticism is at the center of an ongoing and heated debate. From a utilitarian perspective, the debate consists of a weighing of the costs versus the benefits of using animals for research. On the one hand, there is the pain and suffering

that the animals experience. On the other, there are the benefits to be gleaned from the research. A partial solution to this quandary is to minimize these costs through the use of more humane treatment of laboratory animals, keeping them in good health, and maintaining them in comfortable, sanitary conditions. Other solutions include the proper administration of anesthetics and analgesics to the animals before and during surgery and preventing infections through the use of antibiotics and proper surgical techniques (Carlson, 2001). In the United States and other countries there are currently oversight agencies in place that have developed regulations governing animal care. However, animal rights activists argue that enforcement of these regulations by the agencies has been less than rigorous. These activists advocate a greater number of surprise inspections and the setting of stricter regulations that have to do with caging, feeding, and exercise.

On the other side of the balance: How is it possible to justify the use of animals for research? The scientific community points out the store of knowledge and the wealth of benefits that have come from the use of animal subjects. These include the discovery of insulin, the prevention of polio and tetanus, and treatments for anthrax and small pox. Many of the medical discoveries that have sprung from studies that used animals have additionally benefited animals themselves—the discovery of treatments for feline leukemia being a prime example. Animal rights activists counter that alternative, noninvasive testing procedures, such as cell culture and computer simulations, should be used instead.

Brain Recording Techniques

In **single-cell recording**, a very fine microelectrode is inserted into either a single neuron or the extracellular fluid adjacent to it. Changes in that cell's electrical conductivity or its rate of firing can then be measured. In **multiple-unit recording**, a larger electrode is used to measure the collective electrical activity of a group of neurons. In a classic study Hubel and Wiesel (1962) recorded activity in single cells of a cat's visual cortex. They discovered that specific cells responded to lines of a specific orientation in the cat's visual field, implying that line orientation constitutes a basic visual feature of pattern recognition. Although cell recording can give us very detailed information about what one neuron or a relatively small number of neurons is doing, it fails to yield the "big picture." That is, it does not inform us about more global brain activity.

If one wants to investigate what the brain as a whole is up to, he or she can choose from a variety of other tools. The **electroencephalogram** (EEG) is a

recording of the brain's gross electrical action. The procedure uses large electrodes that are taped to the scalp. The output of an EEG is in the form of wave patterns. This method has been used extensively in the measurement of brain activity during sleep. From research that has used this method we have learned that the brain passes through several distinct sleep stages, each having its own pattern of wave activity.

The EEG can also be used to measure a subject's brain activity that is in response to his or her experience of a particular event. The resulting waves that are recorded are called **event-related potentials** (ERPs). For instance, a researcher could sound a tone to a participant in a study and then record the subsequently occurring brain activity. This has been used in many studies to yield insights into the brain mechanisms that underlie perception and attention. A problem with the EEG is that it is a very crude measurement. What researchers usually want to know is which specific brain areas become active during a specific type of brain functioning, and in what order. This information is important because it tells us something about how information is processed.

Modern technology has produced new **brain imaging** techniques that allow us to "peer inside" the head with greater accuracy than ever before. These techniques allow imagery of both brain structure and function. They allow us to see not only the static, three-dimensional organization of brain areas, but also the dynamic activities of these areas as they unfold in time. Although there are many brain mapping procedures that have clinical and medical purposes, we will describe here only those three that have the greatest relevance to cognitive science research.

Computerized Axial Tomography (CAT). This technique was first developed in the 1970s. With this technique, an X-ray source rotates around the brain. Low-level X-ray beams pass through the brain and are picked up by a detector positioned on the opposite side. The information from all of the individual X-ray scans is collected and processed by a computer, and two- and three-dimensional views of the underlying tissues are then constructed. A CAT scan typically yields an image of a single cross section of the brain. A downside of this procedure is that it does not have fine temporal resolution. It can only produce images at the rate of one per second, and so may miss very fleeting mental events. New techniques have partly overcome this resolution problem. CAT procedures are best used for structural-level analysis.

Positron Emission Tomography (PET). This imaging procedure (developed in the 1980s) was developed later than computerized axial tomography.

PET scans measure blood flow in the brain while a participant is carrying out a cognitive task. This is accomplished through the use of radioactive isotopes (tracers) attached to carrier molecules, such as glucose or oxygen molecules. The molecules are injected into the participant's bloodstream, whereupon they make their way to the brain. Brain areas that are more active during the carrying out of the task will show greater regional cerebral blood flow and, correspondingly, a greater concentration of the tracer molecules bearing the isotopes. The activity in these areas is measured using a detection device that counts the positron particles that are emitted by the isotopes. PET scans are especially useful for mapping those parts of the brain that are involved in specific cognitive operations, such as visual imagery, language, and memory.

PET scans have a leg up on CAT scans in that they are better at depicting brain function. They have a fairly good spatial resolution, being able to monitor locational changes in brain activity to within just a few millimeters. PET scans lack some of the temporal resolution of other brain mapping techniques, however. They cannot show rapid changes in brain activity—those that occur over time periods that are on the order of milliseconds. The radioactive isotopes that are used are expensive and their radioactivity is short-lived. There is also a small risk associated with the introduction of radioactive compounds into the body.

Magnetic Resonance Imaging (MRI). In most MRI evaluations, a patient is placed inside a tube that contains a powerful magnet. Protons, which are subatomic particles present everywhere in the body, align themselves in the magnetic field in the same way iron filings organize themselves around a small magnet. A radio-wave pulse is then applied to the brain or other part of the body undergoing the scan. The radio signals are bounced back and picked up by a detector unit. The reflected signals exhibit different characteristics that are determined by the nature of the atoms that have been penetrated by the signals and their surrounding chemical environment and are converted to images. The cross-sectional images show the structure of internal soft tissue.

A recently developed variant of this procedure is called **functional magnetic resonance imaging**, or fMRI for short, and is used to reveal changes in brain activity over time. Like PET scans, fMRI scans detect alterations in local blood flow and oxygen level. Brain areas that show increases in these measures are those that have been activated during specific cognitive operations. So, fMRI scans, like PET scans, are used to map out these active brain areas. fMRI scans provide better spatial resolution than CAT scans without any of

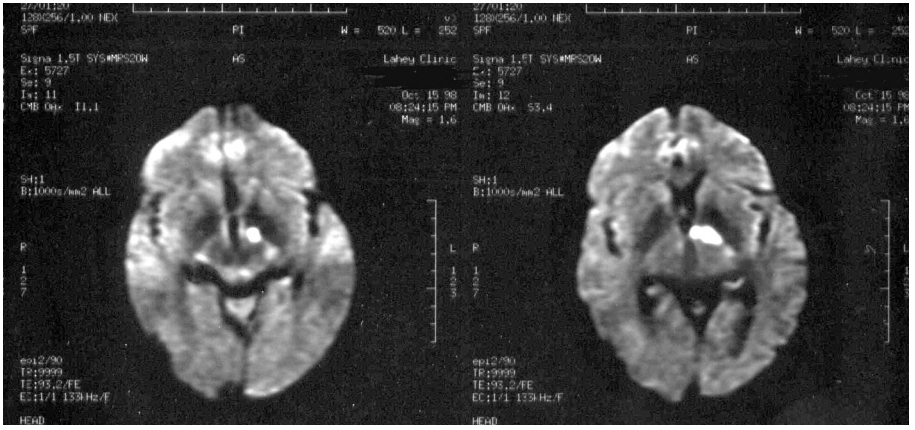


Figure 6.1 Results of an MRI scan showing a two horizontal sections of the brain. The white areas indicate a region of increased activity

the risks associated with the injection of radioactive isotopes. They are the current method of choice for investigation of brain function. Figure 6.1 depicts an MRI image showing a horizontal scan of the brain.

The Electrical Stimulation Technique

So far, we have examined procedures that have to do with brain damage or the recording of brain activity. There is a third procedure that is neither as intrusive nor as passive. This method involves the actual activation of a specific brain area via **electrical stimulation**. An electrical current is passed through a bipolar electrode, which causes the neurons of a localized area of brain tissue to become active. The resulting behavioral effects are usually the opposite of those observed in the brain lesion technique. Lesions prevent neurons from firing, resulting in negative symptoms—the absence of some behavior that is believed to be governed by the affected region. Stimulation, on the other hand, encourages neuronal firing, resulting in positive symptoms—the facilitation of the behavior governed by the area. A difficulty with this procedure is that stimulation might induce supra-normal activity, or an over-activation of the region and the areas associated with it. This would produce behaviors not associated with normal functioning.

The Small Picture: Neuron Anatomy and Physiology

Neurons are the microscopic basis of the brain. They are the individual functional units that perform computations. The purpose of a neuron is to conduct a message in the form of an electrical impulse. A neuron can be thought of as a link in a complex chain because it receives messages from other neurons and then "makes a decision" whether to send a message of its own. Figure 6.2 depicts the major structures of a typical neuron. Messages are received by the feathery projections that are known as **dendrites**. Dendrites form an extensive branching "tree," which connects the neuron to many other neurons. Any incoming messages picked up by the dendrites are then passed along the cell body.

A process whereby the neuron sums up all the inputs it receives from other neurons determines whether or not it will fire. The cell fires if the sum of these inputs exceeds the cell's **threshold of excitation**. This process represents a sufficient change in the neuron's resting electrical state. If the "decision" to fire is made, an electrical signal called an **action potential** is initiated. The action potential then propagates down the **axon**, a long tubular structure that projects outward from the cell body. The axon, which can extend for some distance, ends in a **terminal button**.

The terminal button does not have direct physical contact with the next cell. Instead, there is a gap between the two cells that is known as the **synaptic cleft** (see Figure 6.3). How, then, is a message passed from one cell to the next? The answer lies in molecules known as **neurotransmitters**. The job of these neurotransmitters is to complete the transmission of the signal across the synapse. When the action potential arrives at the terminal button, it triggers the release of neurotransmitter molecules into the synapse. The transmitters diffuse across the synaptic cleft and attach to receptor molecules located on the dendritic surface of the next cell. A neurotransmitter molecule fits into a cavity in the receptor molecule and activates it, in much the same way that a key fits into and opens up a lock. These activated receptors then contribute to the formation of a new signal in the dendrite of the second cell.

The Big Picture: Brain Anatomy

We now step back a bit and look at the brain from a broader perspective. The brain is a complex structure, with many areas and parts that are associated with particular functions. Rather than catalog all of these, we will describe only the ones relevant to important cognitive processes. There are a number of

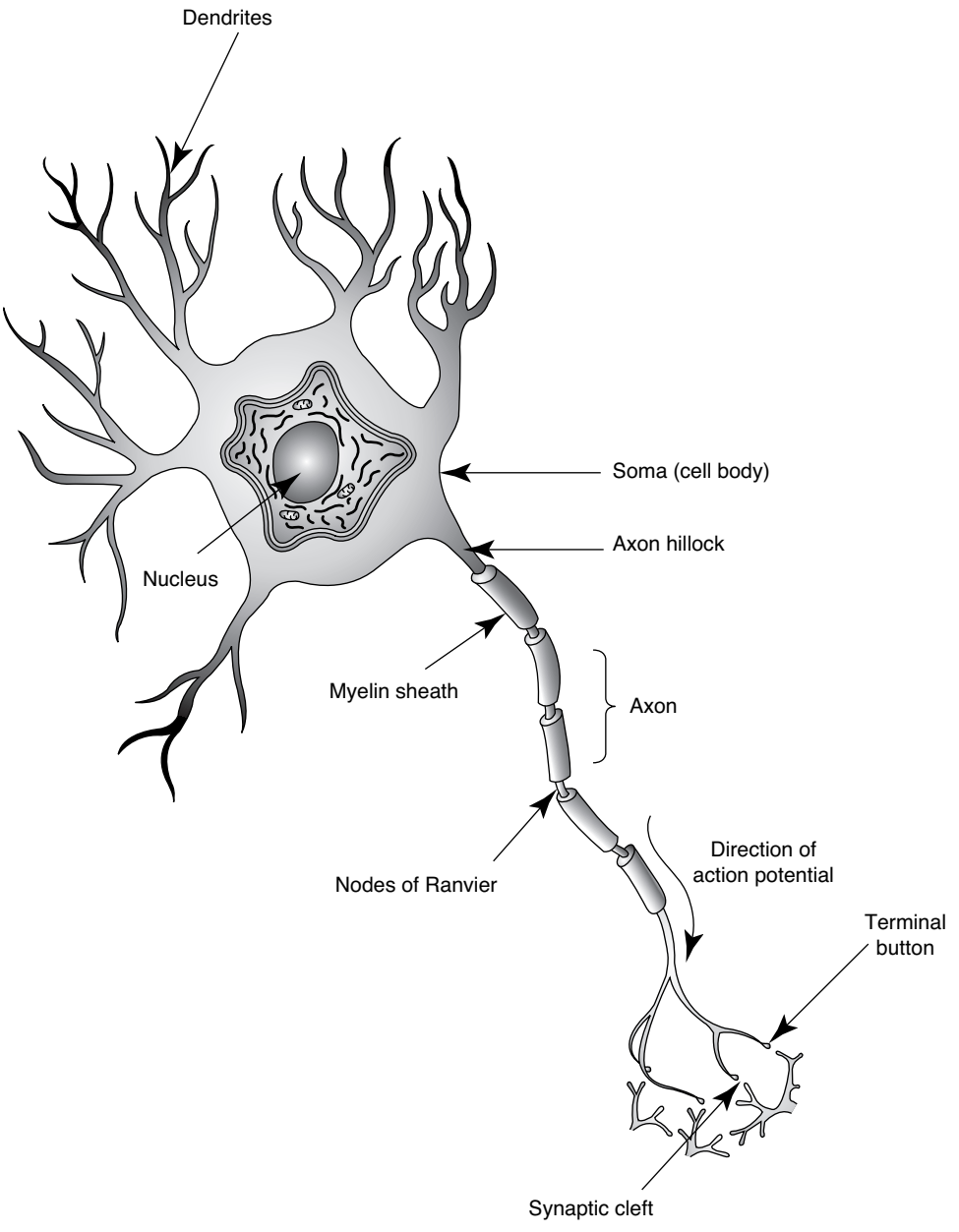


Figure 6.2 Anatomy of a neuron

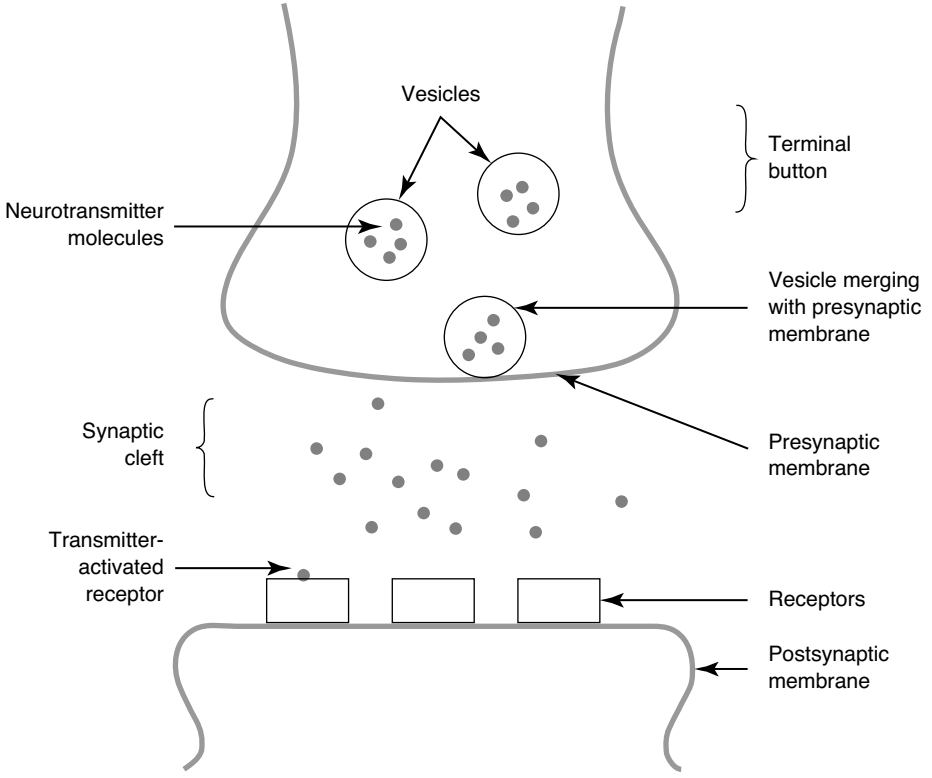


Figure 6.3 Cross-section of a synapse

good texts on this topic that are more comprehensive and we again refer those interested in finding out more to these references.

Directions in the Nervous System

Our discussion of brain anatomy must begin with a discussion of the terms used to imply direction in the nervous system. Anatomists use a special set of terms when referring to these directions. With respect to the human brain, **dorsal** means toward the top and **ventral** means toward the bottom. **Anterior** is used to signify regions that are toward the front of the brain and **posterior**, those that are toward the back. Regions that are located toward the middle of the brain are **medial**, whereas those near the outside are **lateral**. Figure 6.4 illustrates the application of some of these terms to a view of the cortex.

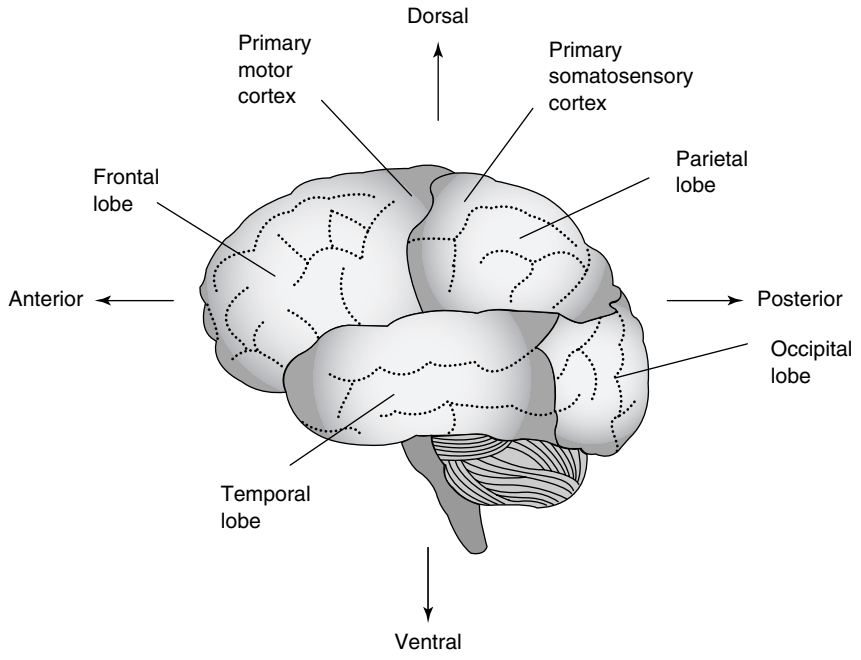


Figure 6.4 Major lobes of the left hemisphere of the human cortex. Some common anatomical directions are also indicated

In addition to the terms that signify these directions, there are also terms that signify imaginary planes that "section" the brain. These planes allow us to consider or to view a "slice" through the brain. There are three main types of planes. A **sagittal plane** "cuts" vertically through the brain, separating it into left and right halves. A **horizontal plane** cuts horizontally, dividing the brain into dorsal and ventral portions. A **coronal plane** also cuts vertically through the brain, but divides it into an anterior and posterior section. These planes are used in the interpretation of PET and fMRI scans.

The Cortex

The cortex is the part of the brain to have been selected for by evolutionary forces most recently and is responsible for a number of higher-order cognitive activities. The cortex is naturally divided into two halves or **cerebral hemispheres**. Each hemisphere is associated with a unique processing style (Sperry, 1985). The left hemisphere is the seat of more analytic, serial, and logical reasoning. The right hemisphere is the seat of more synthetic, parallel, and

relational thought processes. Language function in most persons is localized to the left hemisphere, whereas spatial ability is usually concentrated in the right hemisphere. Information is transferred back and forth between the hemispheres via connecting fibers that are known, collectively, as the **corpus callosum**. The longitudinal fissure separates the two hemispheres.

A **fissure** is a large cleft or separation between two areas of brain tissue, whereas a sulcus refers to a smaller such separation. A **sulcus** typically separates neighboring ridges or folds of tissue. A single ridge of this kind is known as a **gyrus**. At a smaller scale, brain tissue is made up of nuclei (singular: nucleus), which are collections of cell bodies and tracts. The tracts are collections of nerve cell axons.

Figure 6.4 is a view of the cortical regions of the left hemisphere. The first thing you may notice is that the cortex has several large regions, called **lobes**. The **frontal lobe** is located anteriorly and is bounded by the central sulcus and the lateral fissure. It contributes to problem solving and language production. The **temporal lobe** lies ventral to the lateral fissure. The temporal lobe mediates auditory processing, pattern recognition, and language comprehension. Just posterior to the central sulcus is the **parietal lobe**. It governs aspects of attention and spatial processing. The parietal lobe also plays a role in somatosensory processing and in awareness of one's own body. Finally, there is the **occipital lobe**, where visual information begins to undergo more extensive processing. Two other cortical regions should be noted. Anterior to the central fissure is the precentral gyrus. It is the location of the **primary motor cortex.** It is, in effect, a spatial representation or map of the body's parts. Electrical stimulation of the primary motor cortex at a specific point provokes muscular contraction at the site of the corresponding body part. The function of the primary motor cortex is to initiate behavior via the activation of different muscle groups. Just posterior to the primary motor cortex, on the postcentral sulcus, is the **primary somatosensory cortex**. It is also, in effect, a topological body map. Electrical stimulation of any portion of the primary somatosensory cortex triggers the perception of a sensation coming from the corresponding part of the body. The primary somatosensory cortex processes sensory information arriving from the body surface.

The Split Brain

An important feature of the cortex is that information received from the right or left halves of the body is mapped onto the opposite, or **contralateral**, side of the brain. This means that a stimulus presented in the left visual

field—basically everything to the left of your nose—is projected onto the right hemisphere, whereas a stimulus presented in the right visual field is projected onto the left hemisphere. Information presented to one side of the body is thus not processed on the same (or **ipsilateral**) side of the brain. This contralateral organization is the result of the crossing of fibers originating from sensory neurons that are located on one side of the body over to brain areas of the other side. The same holds true for motor control: the left hemisphere sends commands to the right side of the body while the right hemisphere controls the left side.

Experiments with split-brain patients, in whom the corpus callosum has been surgically severed to prevent the spread of epileptic seizures, demonstrate just how specialized the two brain hemispheres are. In these studies, a patient is presented with the name of an object: it is presented to the left or right visual field and so projects to the right or left hemisphere, respectively. The patient can identify the word either by reading it or grasping the object the word represents from among a set of objects that are hidden from view. The utterance of the word can only be accomplished if the information that has been presented to the patient reaches the left hemisphere, as the left hemisphere is specialized for language. The patient's grabbing hold of the correct object with either hand is a form of identification that must be initiated by the contralateral hemisphere. Either hemisphere with information that has to do with the object can lie behind correct identification through grasping because this course of action relies on tactile information alone.

Imagine that the word "spoon" is presented to a split-brain patient in his or her left visual field. The information projects to the right hemisphere. The patient, in all likelihood, cannot read the word because, in most persons, language ability is lateralized to the left hemisphere. He or she also cannot identify the object with his or her right hand because such an action would be controlled by the left hemisphere, which has no knowledge of the spoon. The participant can however reach out and select a spoon with his or her left hand. Imagine another scenario. The word "spoon" is flashed in the right visual field and so projects to the left hemisphere. The participant can at this point say that it is a spoon, as well as grab a spoon with his or her right hand. Correspondingly, this patient would not be able to identify the spoon with the left hand.

The Neuroscience of Visual Object Recognition

You may recall the different models of pattern recognition that were discussed in Chapter 4 (The Cognitive Approach I). These were all ways of describing how we visually recognize objects. In this section, we examine several clinical

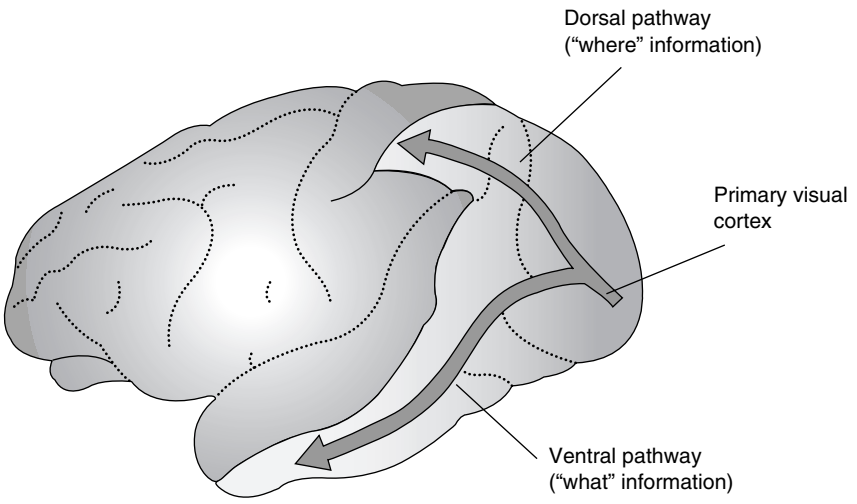


Figure 6.5 The dorsal and ventral pathways of the monkey's visual system

disorders to see what they can tell us about the neurological basis of these recognition mechanisms. But first we must introduce a basic principle of information processing in the visual system—the partitioning of visual input into separate streams.

The visual system, as we mentioned earlier, breaks objects down into their parts or features, with different areas of the brain handling them. Research has revealed two distinct anatomical pathways for the processing of visual information (Ungerleider & Mishkin, 1982). Visual inputs undergo preliminary processing in the primary visual cortex, located in the occipital lobes, at the back of the head. Following this, the information is divided and streams to two geographically diverse parts of the brain. One stream, called the **dorsal visual pathway**, travels upward to the parietal lobe where information about motion and location is extracted. It is sometimes called the "where" pathway because of its representation of the spatial positions of objects. A second stream, the **ventral visual pathway**, carries data about color and form and travels downward to the temporal lobe. It is referred to as the "what" pathway. Figure 6.5 shows the location of both.

Visual Agnosias

Oliver Sacks is a neurologist who is also a gifted writer. In *The Man Who Mistook His Wife for a Hat* he recounts the case histories of several patients

with disorders so bizarre that they border on the fantastical (Sacks, 1985). Take, for example, the case of patient P., who is a music teacher. P.'s gaze is unusual, his eyes dart around as they take in Sacks's face, fixating on one feature and then another. P. is unable to recognize a rose and describes it as "a convoluted red form with a linear green attachment." During one interview with Sacks, P. took hold of his wife's head as he apparently tried to lift it and put it on his own head. Sacks writes: "He had . . . mistaken his wife for a hat!" Although it may seem highly improbable, such patients really do exist. They have a pattern recognition disorder that is called visual agnosia. What is wrong with these patients' brains? How can they make such errors? Let's take a look.

A **visual agnosia** is an inability to recognize a visual object (Farah, 1990). Visual agnosias are perceptual in nature and cannot be attributed to problems that have to do with memory. They are modality-specific. In one case study, a woman who was not able to recognize a picture of a hammer was able to recognize an actual hammer through the use of another sense. She was able to pull a hammer out of a bag that was filled with objects. Furthermore, it seems that most other visual system processing abilities in patients with visual agnosias remain intact. These patients do not have difficulty in detecting motion, interpreting location information, or in recognizing other types of objects. Agnosias are associated with damage to the brain regions that process visual object information.

There are two broad categories of visual agnosia. Persons with **apperceptive agnosia** cannot assemble the parts or features of an object into a meaningful whole. Persons with **associative agnosia** perceive this whole, but have difficulty in assigning a name or label to it. The apperceptive form of the disorder seems to involve a disruption of the formation of an object representation and is therefore more perceptual, or "lower level," in nature. The associative form involves a disruption of the ability to categorize or identify objects, and can be considered more cognitive, or "higher level," in nature.

Apperceptive Agnosia

In patients suffering from apperceptive agnosia, most of their basic visual functioning is intact. This includes the capacity to see details, discriminate between lights of different brightness, and perceive color. They have great difficulty, however, in naming, matching, copying, or telling the difference between simple visual forms. In one test, patients were asked to indicate which of four drawings of common objects matched a single target drawing. The apperceptive agnosic patient Mr. S was unable to do this. He matched a circle to a triangle and a paperclip to a key. Mr. S was also incapable of copying letters of the alphabet. While attempting to copy a capital letter "X," he drew two separated oblique lines.

So how can we explain this disorder? The current explanation is that apperceptive agnosic patients have suffered a disruption of their perceptual grouping mechanism. In the psychological approach chapter we described that grouping is the process of assembling basic perceptual features. These agnosics lack the ability to integrate the various features of an object, although they can perceive the individual features accurately. For example, an apperceptive agnosic patient might be able to recognize that a picture of a car has wheels, windows, and doors, but he or she could not combine these parts into a complete percept of a car. These individuals have sustained general damage to the occipital lobes and nearby areas. In these individuals, this kind of damage interferes with the combination or assembly of features into a unified object representation.

Another disorder that may be considered a type of apperceptive agnosia involves difficulty in recognizing objects when they are viewed from unusual angles or are lit unevenly. This disorder sometimes goes by the name of **perceptual categorization deficit**. It often goes unnoticed, as patients who have the disorder usually have no problems in performing other visual tasks. The existence of the disorder can be revealed under certain testing conditions, however. Affected persons are for instance not able to say that the object represented in a picture of an upright ladder is the same object as that represented in a picture of a ladder as viewed from below. Notice that this deficit constitutes a failure of object constancy, since it is an inability to recognize an object after it has undergone a transformation, such as a change in perspective (see Figure 6.6). These patients have sustained damage to the right hemisphere, especially the right parietal lobe.

Associative Agnosia

Associative agnosias have three diagnostic criteria. First, affected persons have difficulties in recognizing objects visually. Second, these patients can recognize objects using sensory modalities other than vision, for example, touch or sound. Third, they do have the ability to perceive objects holistically, at least in the operational sense of their being able to copy or match drawings of objects. It is with respect to this third criterion that associative agnosics differ from apperceptive agnosics, who cannot perform these tasks.

Individuals with this disorder demonstrate a behavioral anomaly. They can copy a drawing accurately, albeit quite slowly, but cannot name what it is they just copied. One patient, L. H., copied line drawings of a teabag, a diamond ring, and a pen rather precisely, but could not supply the correct verbal label for any of the items (Levine & Calvanio, 1989). This implies that associative agnosics are able to perceive entire objects correctly. Their difficulty lies in coming up with a name for what they see.

Foreshortened view

Minimal feature view

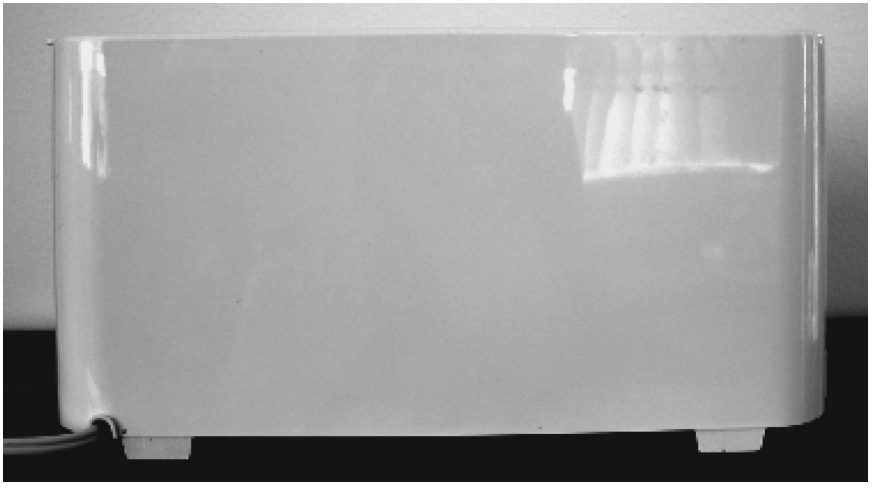


Figure 6.6 Patients with visual agnosia can have difficulty in recognizing an object when it is viewed from an unusual angle (foreshortened view) or from a perspective with few features visible (minimal feature view).

Several theoretical explanations of this disorder have been proposed (Farah, 1990). Early explanations posited a disconnection between language areas that contain linguistic representations and visual areas. Under these theories the two areas have remained intact; it is the connections between them that are damaged, which would preclude the matching of visual inputs to verbal labels. Humphreys and Riddoch (1987) later elaborated upon this idea. They believed that the deficit was owing to damage to a system that contains stored visual object representations, which would prevent the matching of one to the other. More recent explanations use neural network pattern recognition architectures (Hinton, 1981). Damage to the connections that enable the formation of object representations is then used to account for the observed anomalous behaviors.

Face Perception

Prosopagnosia is an inability to recognize faces, despite the capacity to recognize other types of visual stimuli and the presence of generally intact intellectual functioning. Prosopagnosia is considered a type of associative agnosia because, in persons who have the disorder, the ability to perceive faces is mostly intact. The difficulty again lies in recognition: in one's being able to identify the faces. Individuals with this disorder are sometimes unable to recognize close friends, family members, and, in some cases, even the reflection of their own face in the mirror (Burton et al., 1991; Parkin, 1996)! The faculty of discrimination is also impaired. Prosopagnosics have problems in telling faces apart.

Research with monkeys has uncovered the existence of cells that respond selectively to faces. The area that has been identified is in the inferotemporal cortex (IT), part of the ventral pathway that is responsible for the recognition of forms. Bruce et al. (1981) found that some IT neurons fired most rapidly when the animal was presented with the stimulus of a complete face, either that of a monkey or a human being. The neurons fired less rapidly when the monkeys were presented with images of incomplete faces—those that had one feature, such as the eyes, removed. A caricature of a face produced even slower responding. Wachsmuth, Oram, and Perret (1994) also measured neural firing in area IT in monkeys. They discovered that there were cells that became active when the animal was shown a picture of a face or a face and a body. Response dropped significantly when the stimulus consisted of a picture of a body only (with the face covered up).

It is worth digressing for a moment to discuss the idea of neural coding. The research described in the preceding paragraph makes it seem possible that we have individual neurons that respond to individual faces—that there are single or multiple cells that fire only when the stimulus consists of a particular face,

such as that of your grandmother. This type of representation is known as **specificity coding**. In specificity coding, a single cell fires only in response to the presence of a particular face, say, your grandmother's, but not in response to any other face. Another cell would fire only in response to your grandfather's face, but not to your grandmother's or anybody else's. Although specificity coding of this sort is possible, it is unlikely for two reasons (Goldstein, 2002). First, it is now known that cells that respond to one face often respond to others as well. Second, a tremendous number of cells would be required for the coding of all the faces we know, and of all possible views and expressions of these faces.

A more likely explanation is **distributed coding**, in which a specific face is coded for by a specific pattern of activation among a group of cells. In this scheme your grandmother's face might cause cell A to fire rapidly, cell B to fire slowly, and cell C to fire at a moderate rate. Your grandfather's face might induce a different pattern of activation among these same three cells. Distributed coding can efficiently represent a large number of faces or facial attributes and in a way that uses many fewer neurons.

In human beings "face cells" seem to be located in the **fusiform face area** (FFA). Studies that have relied on fMRI show that visual stimuli that consist of pictures of faces activate this area, located in the fusiform gyrus of the human IT (Clark et al., 1996; Puce et al., 1995). This region appears to be dedicated to the processing of face information (Kanwisher et al., 1997). That there should be a part of the brain devoted to faces is not surprising, considering the important role they play in social interaction. Evolutionary forces have probably selected for a face-processing mechanism, although there is an ongoing debate over whether faces constitute a special class of perceptual stimulus.

The Neuroscience of Attention

Attention plays a key role in many different cognitive phenomena. This may explain why there are at least six distinct brain structures that underlie attentional effects (Posner & Peterson, 1990). We can think of these structures as being interconnected and as collectively forming an attentional neural network. Figure 6.7 shows their locations. These different areas work, sometimes in concert, to control attention.

We will describe the location and function of each of these structures in detail in the numbered list that follows, but first let us give you a sense of how they work together. The reticular activating system is responsible for our

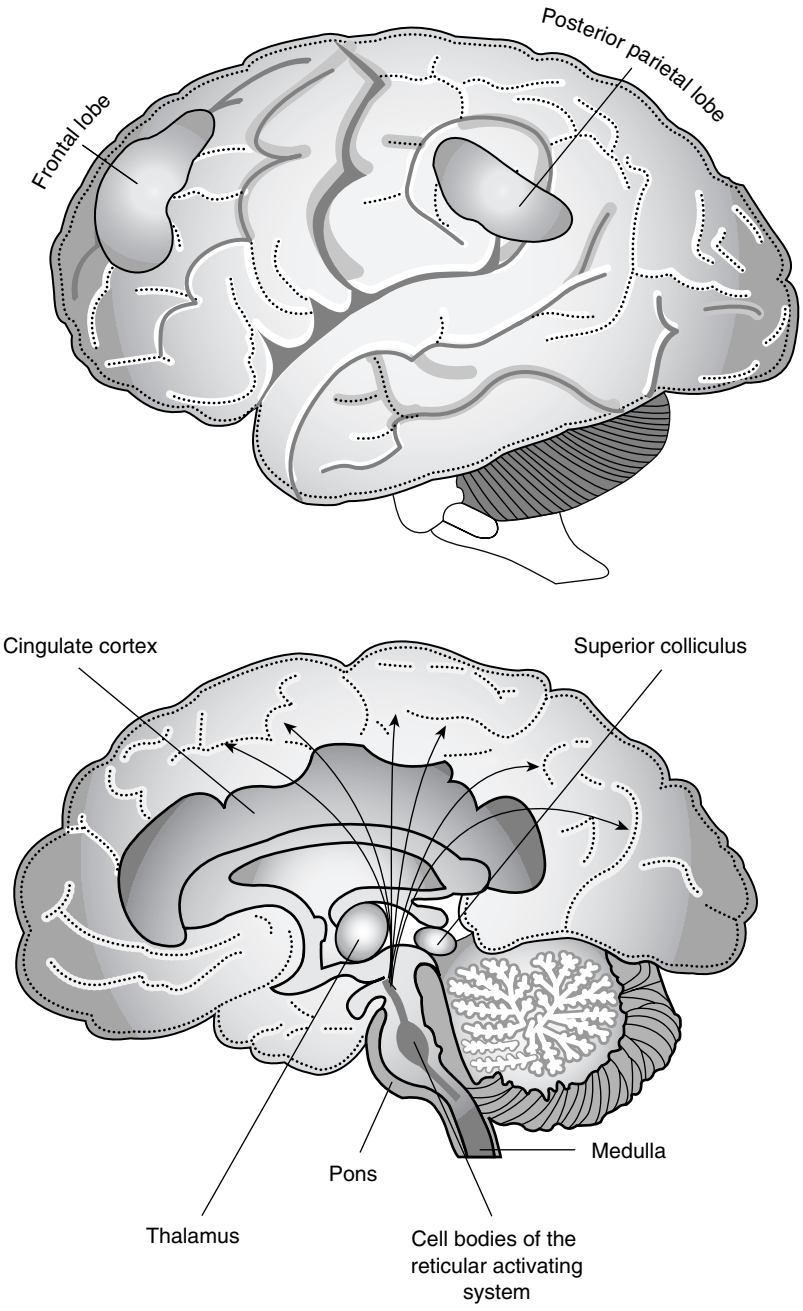


Figure 6.7 Brain structures involved in the function and control of attention

overall arousal level, that is, the extent to which we will pay attention to anything. The thalamus, in conjunction with input to the reticular activating system, regulates the amount of incoming sensory information that reaches the cortex for processing. A more refined processing of the information becomes possible once it reaches the cortex. This processing becomes the job of the parietal lobe, which allocates attentional resources to certain aspects of the information, such as an object's spatial location or some other feature of the object. Ultimately, the cingulate cortex initiates a response on the basis of what is attended.

The structures are:

1. The **reticular activating system** (RAS), located in the midbrain, consists of a network of about 100 nuclei. These nuclei have projections to diverse areas of the cortex. The function of this system is very basic: it controls the brain's overall arousal and alertness levels. Activation of the RAS is linked to the ability to sustain attention over time. Bilateral lesions to the RAS result in a comatose state in which patients are unresponsive to most stimuli.

2. The **superior colliculus** is located in the midbrain, dorsal to the hindbrain. Its function seems to be the shifting of one's visual attention from one object or spot in one's visual field to another. Monkeys in which superior colliculus lesions have been induced are unable to shift their gaze to a novel stimulus that has turned up suddenly in the peripheral visual field (Schiller, Sandel & Maunsell, 1987). This type of deficit has also been observed in humans in whom there has been a degeneration of the colliculus (Rafal et al., 1988).

3. The **thalamus** is dorsal and anterior to the superior colliculus. It receives inputs from the RAS and forwards them to the cortex. It thus is a part of the general RAS arousal circuit. The thalamus also serves as a relay center; it forwards incoming messages that correspond to the different sensory modalities to parts of the cortex specialized for processing them. It is believed that a particular nucleus within the thalamus is responsible for regulating those sensory messages. It serves as a filter or gatekeeper, as it determines how much of this information is allowed to continue to the cortex for further processing. A study that used PET observed a greater amount of thalamic activity in instances in which a participant was asked to single out a target item from a collection of items, as compared to when the participant was asked to

identify a target item that had been presented by itself (LaBerge & Bochsbaum, 1990).

4. The parietal lobe seems to service a broad variety of visual attention-related tasks. In monkeys, cells in the **intraparietal sulcus** of the parietal lobe are active when a particular spatial location is attended (Colby, Duhamel & Goldberg, 1996). In human beings the intraparietal sulcus is active in processing situations that entail the conflation of visual stimulus features such as color and shape, which suggests that it is responsible for the binding together of features in visual search (Wojciulik & Kanwisher, 1999). The parietal lobe is also the site where attentional resources are allocated to different tasks (Coull & Nobre, 1998; Fink et al., 1997). Think back to Kahneman's capacity model of attention from Chapter 4. You may remember that there was a component of that model called the allocation policy. Its job is to distribute attentional resources among a range of possible activities. The parietal lobe appears to be the neural equivalent of that component.

5. The **cingulate cortex** is believed to be the site where a response is selected, especially in instances in which that response entails the inhibition of or the ignoring of an alternative response. An example of the selection of a response in the presence of competing inputs is the Stroop effect (Besner & Stolz, 1999). As an illustration of this well-known phenomenon: participants are asked to look at a word that has been printed in color, and to identify the color or to read the word. The two attributes can be congruent, as when the word "red" has been printed in the color red, or incongruent, as when the word "red" has been printed in the color blue. In the congruent condition, observers react quickly because the two attributes prompt similar responses. In the incongruent condition, observers' reactions are slowed, because one of the attributes elicits a response that must be ignored. Milham et al. (2001) observed an activation of a portion of the cingulate cortex in participants who were required to perform tasks in the incongruent condition.

6. We have already introduced the frontal lobes. With respect to attention-related tasks, their job, like that of the parietal lobes, is quite varied. The frontal lobes play a role in the selection of motor responses, in goal-directed action, and in adherence to a task, among other things. The frontal lobes feature prominently in executive function and problem solving situations, and so we reserve the greater part of our discussion of the frontal lobes for later.

Models of Attention

Baniche (2004) outlines several major categories of attention models. We describe two of them here. In component process models, distinct brain areas each have a unique and non-redundant function. In distributed network models, the areas can be characterized as having some operational overlap.

A Component Process Model

Posner, Inhoff, Friedrich, and Cohen (1987) propose that each of the multiple brain areas responsible for the control of attention performs a distinct operation. Their model specifically describes the changes that occur in visual selective attention, where attention is shifted from one spatial location to another. They posit that cells of the parietal lobe are used to disengage attention or to remove it from a specific location. The superior colliculus then moves attention to a new location. Finally, the thalamus engages the attention and focuses it on the new location.

The data that support this model come from an experimental paradigm in which participants are asked to focus on a central cross and are then cued to respond to a target, such as an asterisk, that can appear to either side of the cross (Posner, 1980). The target appears in one of two boxes, to the left or right of the cross. There are two types of trials. In "valid" trials, the box that will contain the target lights up. In "invalid" trials, the box that will not contain the target lights up. In conditions in which the majority of the trials are valid, participants are faster to respond during valid trials, because the cue is a reliable indicator of where the target will appear.

In a variation of the Posner visual selective attention paradigm, patients with damage to either the right or left parietal lobe had no difficulty responding when the cue and target appeared on the side contralateral to the damage. In these cases the cue was valid and the damaged parietal lobe, which processed information coming from the opposite visual field, did not have to disengage attention. But during invalid trials in which the cue appeared on the side ipsilateral to the lesion and the target appeared on the other side, response time was slowed. In these cases the parietal lobe of the damaged hemisphere did have to disengage attention (Posner, Walker, Friedrich & Rafal, 1984).

Patients with collicular damage show delayed responses in valid trials. This suggests they are slower to shift attention from the fixation cross to the cued target location, no matter which side of the cross it is on (Rafal et al., 1988). Individuals that have sustained damage to the thalamus exhibit yet another result. They are slower to respond whenever the target appears on the side

contralateral to the side of the thalamus that is damaged, whether the trial is valid or invalid. This implies that they cannot engage attention upon the target.

Distributed Network Models

Mesulam (1981) proposes an alternative model for the neural machinery that controls attention. In this model, the separate neural structures are not specialized and functionally independent, as they are in the component process model. Instead, the functions of the different areas overlap to some degree. Each brain region performs a major operation that is attention-related but can perform other attention-related functions as well. This redundancy in the network implies that any given area can suffer damage while the system as a whole will maintain some of the functionality subsumed by the damaged region.

Mesulam's (1981) model explains how attention is directed to extrapersonal space, that is, to areas outside the body, and is derived in part from the patterns of deficits that are observed in connection with specific types of brain damage, in monkeys and in humans. The model encompasses four brain areas, each playing a primary but not exclusive functional role in the control of an aspect of attention. First, the posterior parietal cortex provides a sensory map or representation of the space in which attention will be directed. Second, the cingulate gyrus in the limbic cortex plays a motivational role. It determines what should be paid attention to and what can be ignored. Third, the frontal cortex coordinates motor programs for attention-related actions. These actions would include fixating on certain regions in the visual field, scanning across the visual field, or reaching out to grasp an object. Finally, reticular structures generate arousal and vigilance levels.

Let's give an example that will illustrate how all these areas might work together. Imagine that Susan is sitting at the breakfast table. She missed dinner the night before and so is quite hungry. Her posterior parietal cortex would contain a sensory representation of the breakfast table that included the locations of items such as a glass of orange juice and a bowl of cereal. Her cingulate cortex would direct her attention to items and/or regions of significance. In this case she would pay more attention to the orange juice than to the tablecloth. Her frontal cortex would supervise motor actions, such as her looking at the glass of juice and reaching out to pick it up. Underlying all this would be arousal levels that have been determined by her reticular system. If Susan had a good night's rest, her arousal and vigilance ought to be at high levels.

In some versions of distributed network models, larger brain regions process attention differently from what is outlined above. Posner (1992) breaks up

the system into posterior and anterior attentional networks. The first network consists of structures located posteriorly and is concerned primarily with the selection of information on the basis of the sensory characteristics of external stimuli. The second network is made up of frontally located structures, primarily, and selects information on the basis of abstract representations, such as meaning. It is important to emphasize that these areas are not mutually exclusive in the performance of their operations; there is some shared functionality (Posner, Sandson, Dhawan & Shulman, 1989).

If multiple brain areas are responsible for our attention, then you might be wondering how these areas coordinate their activity. This might be achieved via a synchronization of nerve cell firing in each area. To learn more about how such a synchronization might take place in perception, see the In Depth section.

The Neuroscience of Memory

Early neuroscience research on memory asked a very basic question: Where in the brain are memories located? Karl Lashley (1950) attempted to find out. He was searching for the **engram**, a physical change in a part of the brain that is associated with learning. His methodology was to train monkeys to perform different tasks, for example, to open up a box that has a latch, in the expectation that the animals would form a memory of the task. He would then introduce lesions to parts of the brain to see if the memory of the task was destroyed in the process. Lashley's rationale was that if you destroy the part of the brain that contains the engram for the task, the monkey's memory of how to perform it will have been lost. In one experiment, he destroyed larger and larger areas of cortex in a graded manner. He found that the greater the amount of tissue that had been destroyed, the greater was the amount of time needed to retrain the monkeys to perform the task. But in no case did the animals actually forget how to perform the task.

Lashley concluded that memories are not laid down in any one area of the brain, but that all parts of the brain participate in memory storage—a principle referred to as **equipotentiality**. This principle has some interesting implications. It suggests that multiple copies of a single memory are stored throughout the brain and that multiple brain regions participate in memory formation. We now know that there is some truth to this statement. The distinct cortical areas that process information derived from a single sensory modality do contain very different representations. It is also the case that many brain areas are

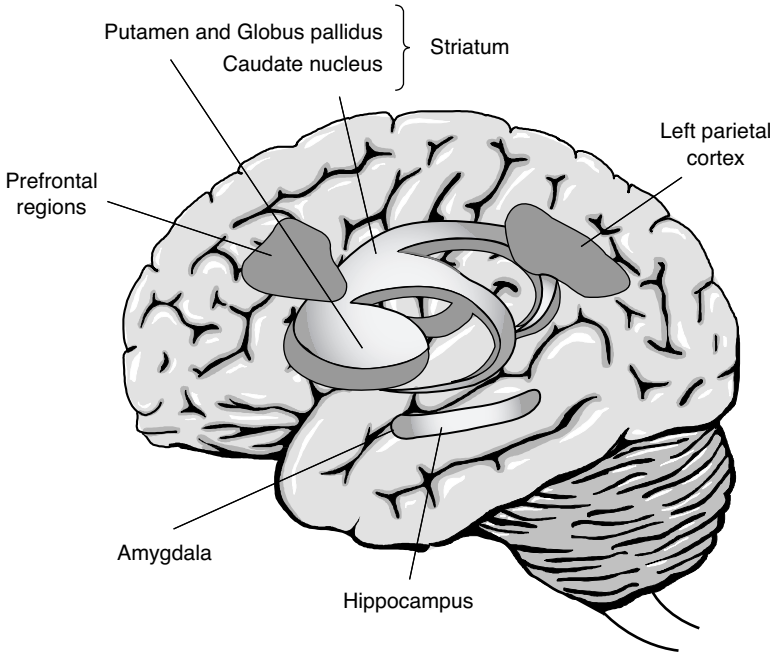


Figure 6.8 Brain areas that form part of the human memory system

involved in memory. These areas include the hippocampus, the entorhinal cortex, the amygdala, the striatum, the left parietal cortex, and the prefrontal regions (see Figure 6.8). In this section we will examine some of the more recent work on the neuroscience of memory. Our focus, as before, is to describe the different brain structures that underlie memory processes and to formulate models of how they work together to create memories.

Learning and Memory

Learning has taken place when an event has provoked a nervous system alteration that in turn provokes a change in the organism's behavior. It is this change in the nervous system that is a memory. From a neuroscience point of view, a memory is essentially the formation of a neural circuit—the forging of a pathway among neurons. The memory is then contained within the circuit. When the circuit is reactivated at some point in time after its establishment,

the memory that corresponds to the circuit is experienced and informs the organism's behavior. In this way, nervous systems preserve the past and help organisms to adapt to their environments.

Learning requires **synaptic plasticity**, which refers to a capacity for change in the structure or biochemistry of a synapse. Donald Hebb (1949) was perhaps the first to understand this and formulated what has become known as the **Hebb rule**. It states that if two connected neurons are active simultaneously, the synapse between them will become strengthened. Take two cells, A and B. That there has been an increase in the strength of their connection means that when cell A is active, it is more likely to activate cell B than it had been formerly. Cells A and B have now become linked together in a circuit: when A fires, B fires also.

The specific neural change that mediates this Hebbian form of learning is called **long-term potentiation** (LTP). It is the enduring facilitation of synaptic transmission that occurs following activation of a synapse by intense high-frequency stimulation. Rapid repeated stimulation is the key to LTP. It causes a buildup of electrical activity in the cell and induces changes, such as an increase in the number of receptors (Tocco, Maren, Shors, Baudry & Thompson, 1992) as well as other structural changes that result in increased neurotransmitter release (Edwards, 1995). These factors strengthen the synapse, increasing its responsiveness to further stimulation. LTP occurs in parts of the brain where learning and the creation of new neural circuits are important. It has been most often studied in the cells of the hippocampal system, which we will now discuss.

The Hippocampal System

We begin our discussion of the hippocampus with a famous case study (Corkin, 1984). H. M. was a patient who suffered from **epilepsy**, a disorder in which neurons fire uncontrollably, producing muscle spasms and seizures. H. M. did not respond to anticonvulsant medication. To stop his seizures, which were severe and debilitating, surgeons took drastic measures. They removed portions of his medial temporal lobes, including the hippocampus. The surgery was successful in the sense that it put an end to his seizures, but it left him with a profound memory deficit. After the operation H. M. lacked the ability to learn any new information. He could have a conversation with a friend, turn around to talk to someone else and, in just a minute or so, completely forget that his friend was even there. The removal of the medial temporal lobe

tissue disrupted the process by which information is transferred from working or short-term memory to long-term memory—called **consolidation**.

Despite his memory deficit, many of H. M.'s other memory abilities were unaffected. His working memory and problem solving capacity were intact. He in fact enjoyed solving crossword puzzles—but would sometimes solve the same puzzle several times, as he forgot that he had already completed them! The content of his long-term memory prior to the surgery was also unaffected. This was true for both declarative and procedural long-term memory. He could remember the location of the house he lived in and could still perform the skills he had acquired prior to the surgery. The case of H. M. demonstrates that different memory functions are handled by different neural structures. Damage to one of those structures can selectively remove a component of memory ability. This is a theme that is repeated later in this section.

Hippocampal damage is associated with **anterograde amnesia**, an inability to retain new information following some traumatic incident. H. M. suffered from this type of amnesia, as he could not retain new information after the surgery. This form of amnesia needs to be distinguished from **retrograde amnesia**, an inability to remember information acquired prior to the traumatic event. Retrograde amnesia is often caused by head injury and in most cases the loss extends only as far back as a week or so before the injury.

It is the hippocampus and associated areas that are responsible for the consolidation of newly acquired information. Before we can understand how the hippocampus functions, we need to familiarize ourselves with its rather complex anatomy. The hippocampus is just one of a collection of structures in the limbic system. Figure 6.9 is a diagram of the circuits that run through the hippocampal system.

With these pictures in mind, let's take a guided tour of this amazing area. Sensory information from the olfactory, frontal, parietal, and temporal lobes of the cortex is first passed to the superficial layers of the entorhinal cortex. From there, it is transmitted via the perforant path to cells in the dentate gyrus. These cells then send signals through mossy fibers to another collection of cells in field CA3, which in turn send projections by Schaffer collaterals to field CA1. The axons of CA1 cells feed to the subicular complex and then return to the deep layers of the entorhinal cortex. At this point, information is passed back out to the cortical lobes, forming a loop.

Models of Hippocampal Function

The job of the hippocampus is essentially to take information about "current events" that it has received from the cortex, process this information, and

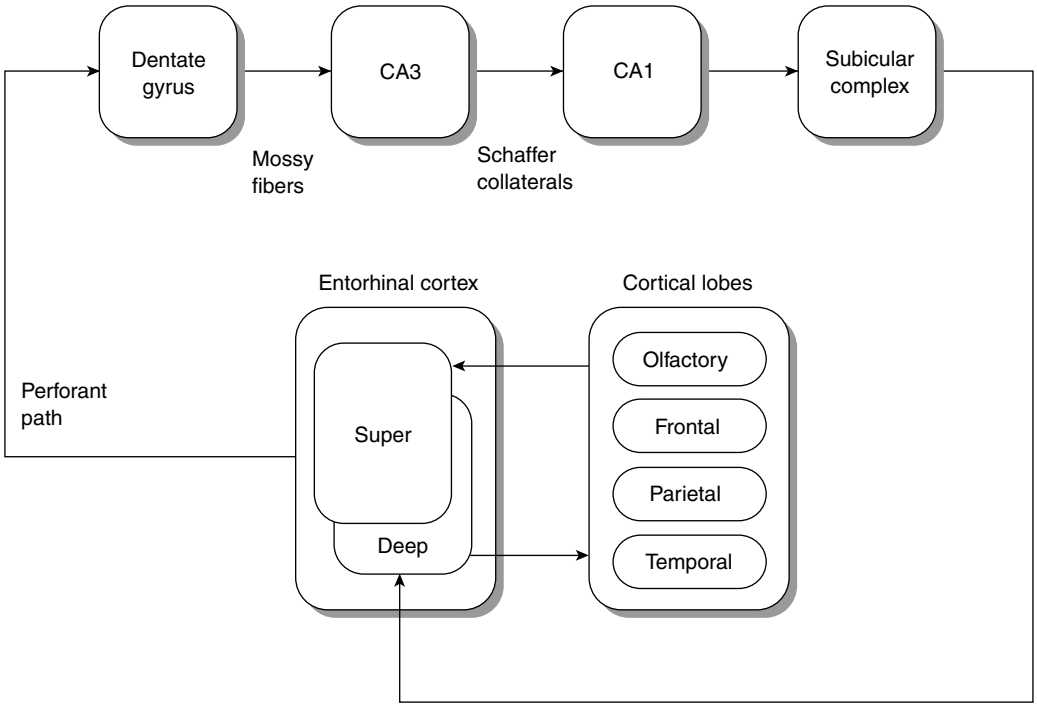


Figure 6.9 Simplified schematic diagram of the hippocampal system

send it back out to the cortex, where it is stored in a more permanent state. Inputs from the cortical regions contain representations of episodes or events that the organism has just experienced. More specifically, we may think of the hippocampus as an integrator. If the job of the cortex is to "divide and conquer" sensory input into distinct processing streams, the job of the hippocampus is to "combine and reunite" them. The hippocampus weaves the disparate parts of an experience into a single unified memory. Let's use the example of a birthday party to illustrate this. The various parts of the cortex have divided up your experience of the party into fragments. The visual system maintains an image of the presents; the auditory system, the sounds of your friend's singing; the gustatory system, representations of the taste of the cake, and so on. All of these representations converge as inputs to the hippocampus, which combines them to form a whole memory of the birthday party. To get

an idea of exactly how it does this, we need to look at computational models of hippocampal function.

There are a number of models that have been designed to simulate the operation of the hippocampal system. These are computational models that use artificial neural networks and that are informed by what is known of the anatomy of the region. Marr (1971) proposed one of the earliest models. His ideas were subsequently tested, and his revised ideas were used to create a newer model (Willshaw & Buckingham, 1990). There are more recently devised models as well (Alvarez & Squire, 1994; McClelland, McNaughton & O'Reilly, 1994; O'Reilly & McClelland, 1994). We will describe the model by Treves and Rolls (1994), because it is more tightly constrained by hippocampal anatomy and physiology than the others.

In the Treves and Rolls model, a neural network in the dentate gyrus stores patterns of entorhinal input that represent fragments of an episode. The network then "teaches" these patterns to the neurons in field CA3. Different groups of neurons in field CA3 code for different scene fragments. The field CA3 to field CA1 projection then combines these fragments into conjunctions that represent larger parts of the entire scene. According to the model, entire episodes are stored in field CA1 for up to several months—forming a sort of "medium-term" memory. Some of these episodes can then be selected for transfer to long-term storage in the cortex. These patterns of activation are fed back out to the cortex through the subicular complex and entorhinal cortex. The outputs then encode the patterns into cortical neurons.

Although we have omitted many of the details, the work by Treves and Rolls (1994) is an excellent example of model building in cognitive science. It uses the methodology of the network approach and "marries" it to some of the hard facts of neuroscience. A number of network models of cognitive function tend to ignore anatomical data. They simulate cognitive operations through the use of standard network architectures, because they are available and because they are well understood. These standard architectures are useful. They can be quite informative and in some cases can lead to insightful computational solutions to problems. But if the goal is to discover how brains work, then a greater effort is needed in the way of constraining the formulation of network models to neuroanatomy.

Neural Substrates of Working Memory

In Chapter 5 (The Cognitive Approach II), we discussed the components of working memory. These were the articulatory loop, where operations on

verbal representations are performed; the visuo-spatial sketchpad, where visual information is processed; and the executive control system, which coordinates activities. Recent neuropsychological evidence confirms the existence of these structures as functionally distinct units and has pinpointed their locations in the brain.

Smith, Jonides, and Koeppe (1996), using PET in an experimental scenario, have discovered the existence of anatomically separate areas for storage and rehearsal of verbal information in working memory. They employed a design in which participants in one condition viewed a stream of letters and were asked to judge whether a given letter matched the one that they had viewed two letters back in the sequence. This task requires both storage and rehearsal. The performance of these participants was compared to that of two control groups. Participants in the first control group were asked to search for a single target letter. This ostensibly requires storage only. Participants in the second control group viewed the letter stream and were asked to make only manual responses to the presentation of a letter as they rehearsed that letter until the next one appeared. This condition requires rehearsal only.

The PET images that were obtained under these study conditions were compared to each other using a subtraction technique. In this technique, the record of brain activity obtained in one condition is subtracted from that of another. According to this logic, if the "two-back" condition calls on storage and rehearsal and the "search" condition calls on storage, the subtraction of the latter from the former should identify the brain areas that are responsible for rehearsal. Similarly, the subtraction of the record of brain activity of the "manual" group from that of the two-back group should identify those brain areas that are involved in storage.

Using this type of analysis, the researchers found that activation of the left hemisphere's posterior parietal cortex corresponded to the storage of verbal material. Three other sites also turned up. All three were in the prefrontal cortex and corresponded to rehearsal for verbal materials. These sites were the inferior frontal gyrus (called Broca's area), another region in the premotor cortex, and a region of the supplementary motor area. These areas are believed to generate a code for use in explicit speech, as we discuss in the linguistics chapter. This code can be used to represent verbal materials for rehearsal and implicit speech as well. These two areas, the left hemisphere's posterior parietal cortex and the left hemisphere's prefrontal cortex can be considered to form part of the articulatory loop that is part of Baddeley's (1986) model of working memory.

In a separate experiment, Smith et al. (1996) investigated the neural substrate of spatial working memory. Participants in this study were presented

with letters that appeared at different positions around an imaginary circle. They were asked to judge whether a given letter appeared in the same position as that of a letter that had appeared three letters back in the series. The resulting PET analysis showed activation in the posterior parietal cortex, but this time in the right hemisphere. The researchers identified this region as the area where spatial information in working memory is stored. They noted that these data were in accord with that of other studies that have identified the same area as the site of storage of spatial information (Jonides, Smith, Koeppe, Awh, Minoshima & Mintun, 1993; Petrides, Alivisatos, Meyer & Evans, 1993).

Evidence from animal studies suggests a location for the rehearsal or maintenance of a spatial code in working memory. Goldman-Rakic (1993) tested monkeys by giving them a delayed response task. In this procedure a monkey is shown two locations, only one of which contains a food item. There is then a forced delay of several seconds, during which the locations are visually obscured by a barrier. The delay forces the monkey to maintain the spatial location of the food item in working memory. The barrier is then removed and the monkey must choose the location that holds the food item in order to get the item as a reward. Monkeys in which the dorsolateral prefrontal cortex has been lesioned cannot perform this task—implying that this region is the site for the maintenance of spatial information in the working memory system.

Studies with both animal and human subjects reveal a third working memory system for objects. This system codes and stores visual object representations. Whereas spatial working memory would code for the location of a letter in space, for example, whether it appears to the left or to the right, visual object memory would code for a visual representation or image of the letter itself, which would include such attributes as its shape and color. This differentiation reflects the existences of the dorsal "where" path for location and the ventral "what" path for identity. Data from Wilson, O'Scalaidhe, and Goldman-Rakic (1993) show that, in monkeys, one brain region mediates object shape memory and another mediates object location memory. The principal sulcus is the area corresponding to spatial location; the inferior convexity, shape location. Both are located in the frontal area. In humans, distinct brain areas lateralized to either hemisphere carry out these functions. Occipital, parietal, and frontal sites in the right hemisphere are linked to spatial memory, whereas parietal and IT sites in the left hemisphere are linked to object memory (Smith & Jonides, 1994; Smith, Jonides, Koeppe, Awh, Schumacher & Minoshima, 1995).

Evaluating the Neuroscience of Working Memory

Jonides and Smith (1997) make a few general conclusions concerning the neural architecture for working memory. They speculate that there may be a separate working memory system for each sensory modality—each having its own storage buffer and rehearsal capacity. The visual object memory system described above would serve vision, another system would serve audition, yet another gustation, and so on. Verbal and spatial working memory, however, are modality-independent. They can represent and rehearse information from any modality.

These investigators are also careful to point out that the various memory codes are in the service of other cognitive processes and they give several examples (Jonides and Smith, 1997). One could, in effect, use the spatial code to form a mental map of how to get somewhere and a visual code to mentally compare the shapes of two objects. What is lacking, they argue, is a more abstract conceptual code that could represent semantic items such as words. This type of code would allow for additional cognitive tasks such as language comprehension, problem solving, and deductive reasoning. There are hints in the literature of the existence of a semantic code of this nature. Future research would be needed to determine whether such a code would have its own attendant working memory structures for storage and rehearsal.

The studies of the neural basis of working memory that we have surveyed thus far have been very informative. They show us that there is in fact a large number of distinct working memory systems that correspond to different sensory modalities and various abstract characteristics. This work can help us to reconceptualize Baddeley's (1986) model of working memory, which originally postulated the existence of verbal and visual components only. We could update this model by incorporating these additional working memory systems. You might recall that Baddeley's model also included an executive control system. This system is responsible for the control of various aspects of working memory function. Executive processes are also used in reasoning and problem solving. For this reason we will talk about reasoning and problem solving in the section on executive function.

Neural Substrates of Long-Term Memories

We have already reviewed some aspects of long-term memory in our discussion of learning and the hippocampal system. Remember that when

information first comes into the brain, multiple areas of the cortex, each corresponding to a unique sensory modality, process it. These representations, however, are short-lived, and would fade away quickly if it were not for the hippocampal system, whose job it is to integrate the disparate aspects of a memory experience and to return the consolidated memory to the cortex. So the cortex is the site where some of our long-term memories reside. But there are multiple forms of long-term memory. Procedural memories store procedural or skill knowledge and are demonstrated through action. Declarative memories store factual or event knowledge and are demonstrated through explanation. The two types of declarative memory are semantic memory, for facts, and episodic memory, for events. See Chapter 5 for a more extensive treatment of these types.

Given this variety of types of long-term memory, what can we say about their neural bases? Are there different brain areas that mediate processing with respect to the different types of long-term memory? The answer is yes. Research in this area shows that they are governed by different brain regions.

Declarative memories rely on the cortex for storage and the hippocampal system for consolidation. But semantic and episodic information are mediated by two separate aspects of this system. This conclusion comes from an examination of case study data. Vargha-Khadem, Gadian, Watkins, Connelly, Van Paesschen, and Mishkin (1997) reported on patients who sustained damage to the hippocampus only early in life. There was in these patients no damage to the limbic cortex of the medial temporal lobe. The limbic cortex consists of the parahippocampal, entorhinal, and perirhinal cortexes. These areas, as we discussed earlier, are considered part of the overall hippocampal system and mediate connections between it and the overall cortex. These patients could not recall anything that happened to them during the course of a day. If they went to a movie or visited a friend they would have no memory of it. They did, however, have excellent memory for facts and did relatively well in school. This suggests that the hippocampus proper is responsible for the consolidation of episodic memory, and that the limbic cortex mediates semantic memory. These researchers also concluded that if both the hippocampus and the limbic cortex were destroyed, all declarative memory ability would be lost. More research is needed to confirm these interesting conclusions.

Learning in procedural memory corresponds to a change in the neural systems that underlie the acquisition of a given task. It is the motor regions that govern skill performance that contain procedural knowledge, not the widespread sensory cortical areas that hold declarative information. Also, procedural memory is not dependent on the hippocampus for consolidation. The **basal ganglia** play a critical role in skill learning (Graybiel, Aosaki,

Flaherty & Kimura, 1994). They are a cluster of brain areas involved in voluntary motor responses. The basal ganglia consist of the striatum, which is made up of the caudate nucleus and the putamen, and the globus pallidus (see Figure 6.8). Changes to the motor cortex have also been documented as taking place during skill learning (Grafton, Mazziotta, Presty, Friston, Frackowiak & Phelps, 1992; Grafton, Woods & Tyszka, 1994). In one study, monkeys trained to pick up small objects such that fine finger and hand movements were required evidenced increased representation of those body areas in the primary motor cortex (Nudo, Milliken, Jenkins & Merzenich 1996).

The Neuroscience of Executive Function and Problem Solving

Executive function refers to cognitive operations such as planning, the sequencing of behavior, the flexible use of information, and goal attainment. Many of these same operations are called upon in problem solving. As we saw in our examination of the cognitive approach, problem solving consists of trying to attain a final goal, which is the problem's solution, via the performance of a sequence of operations that leads to the attainment of individual subgoals. In this section, we discuss the neural basis of such problem solving ability.

The hallmark of frontal lobe damage is a diminished capacity to perform goal-directed behaviors. Patients who have sustained this type of brain damage suffer from **executive dysfunction**, a disorder characterized by a broad array of deficits. Some of these patients have difficulties in initiating actions or in terminating them once they have been initiated. For example, the patients may be listless, may sit around the house all day, and so on, but once they are engaged in some action, such as brushing their teeth, they may not be able to stop. This kind of behavioral deficit is referred to as **psychological inertia**. These patients sometimes appear to be impelled to perform actions that are "suggested by" the environment, such as picking up a pencil and writing with it as soon as they see it. This phenomenon, in which a stimulus in the environment triggers an automatic behavior, is called **environmental dependency syndrome**. Needless to say, these individuals have difficulty in solving even simple problems.

Sequencing, which is the sequential ordering of actions, is necessary in problem solving. Any strategy that is part of problem solving includes sequencing, because multiple steps are simply part of the problem solving process. The Tower of London problem is a tool that researchers have used to test this capacity (Shallice, 1982). Figure 6.10 shows several Tower of London configurations. One sees three sticks of different heights. Three balls of different

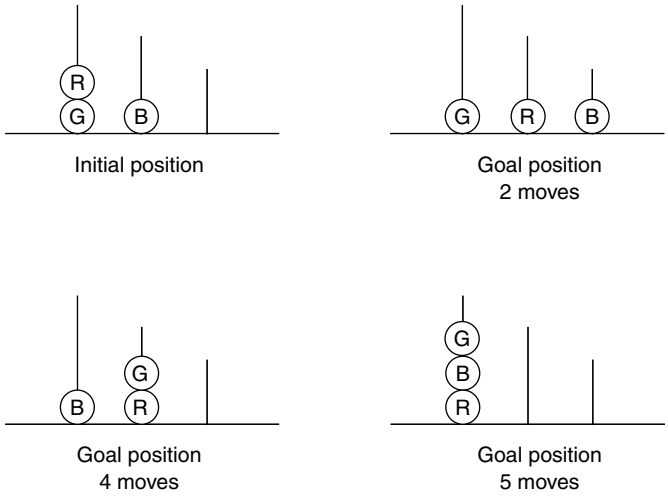


Figure 6.10 The Tower of London task (adapted from Shallice, 1982)

colors have holes through them and can be placed on the sticks. The left-most stick is the tallest and can accommodate all three balls. The middle stick is intermediate in height and can accommodate two balls. The stick to the right is the shortest and can hold only a single ball. The solution to any of the Tower of London problems involves moving the balls that make up an initial starting configuration until they conform to a final target configuration in the smallest number of moves possible.

Patients who have sustained frontal lobe damage, especially when it is to the left frontal lobe, have great difficulty in solving Tower of London problems. Shallice (1982) studied 61 patients with and without localized lesions of either frontal lobe. The patients were then given Tower of London problems. The numbers of these problems that the patients were able to solve in one minute were tallied. The results showed that the group with left anterior frontal lobe damage had scores that were significantly worse than those of the other groups. He concludes that this area is responsible for planning and sequencing of the type that is needed in Tower of London tasks.

In a more recent study, healthy individuals were asked to solve a computer version of the Tower of London task; regional cerebral blood flow (rCBF) in these individuals was measured simultaneously (Morris, Ahmed, Syed & Toone, 1993). Participants who performed the computer task showed increased levels of rCBF in the left prefrontal cortex. Those subjects who took more time to plan their moves and computed the solutions in fewer moves also had greater

rCBF, suggesting that activity in this brain area is responsible for planning during problem solving.

When we examine the insights that are to be gleaned from studies of frontal lobe damage, we see a good example of convergence between cognitive science approaches. In this case the convergence is between neuroscience, on the one hand, and studies of problem solving in the areas of cognition and artificial intelligence on the other. Patients with frontal lobe damage clearly demonstrate deficits that have to do with directing themselves toward a goal or in breaking a problem down into subgoals. They also have difficulty in sequencing their actions appropriately. These components of problem solving are exactly what we said would be required in the solution of any problem, and what computer programs such as GPS and SOAR do to solve problems. The fact that these ideas come up independently in different disciplines shows they are fundamental to an understanding of problem solving.

Theories of Executive Function

In this section we consider two theories of executive function. The theories are cognitive in nature, but have specific neurological implications. Both are formulated to account for automatic versus controlled forms of attention. **Automatic attentional processes** do not require conscious control. They occur when one is experiencing familiar situations. Automatic responses are then triggered by environmental stimuli. Driving a car is an example of an automatic process. It can be carried out virtually without thinking while one executes other actions, such as carrying on a conversation. **Controlled attentional processes**, on the other hand, require conscious control. They are operational when one responds to novel or difficult situations for which there is no prior learned reaction. In these instances, attention must be voluntarily directed. Problem solving calls on this type of processing.

Norman and Shallice (1980) have proposed the first model. It relies on **action schemas**, structures that control automatic attentional processes. In this model, a schema can be triggered by perceptual input or by the output of other schemas. An example of the former would be the image of your unmade bed in the morning. The sight of it would activate a bed-making schema. An action schema can be thought of as an if-then production rule, where perceptual inputs serve as the preconditions necessary to trigger a context-specific action.

In a more general sense, a schema can be thought of as a stored framework or body of knowledge on the subject of some topic (Schank & Ableson, 1977). This type of schema is sometimes referred to as a **script**. The most often-used example is a script for eating in a restaurant. It would consist of a sequence of

events: being seated, getting menus, ordering, waiting, eating, paying the bill, and so on. Scripts can be used for solving problems because they specify goals and the sequences of actions needed to achieve them. We discuss some of the ways in which scripts are used to solve problems in the upcoming chapter on artificial intelligence.

In the Norman and Shallice (1980) model, schemas can be activated independently of one another. This can pose a problem, since, in everyday situations, usually just one or a small number of actions need be performed at any given time. The model specifies two ways in which a subset of appropriate schemas can become activated. The first is **contention scheduling** and is used to govern routine, habitual performances as well as the performance of non-routine, novel tasks. The second selection mechanism is the **Supervisory Attentional System** (SAS) and is used only for non-routine actions.

In contention scheduling, each action schema is executed when the level of activation in one of its control units reaches a certain threshold. However, each schema has mutually inhibitory connections to others, so that the more one schema is activated, the more it suppresses the activation of those it is connected to. This prevents the simultaneous activation of multiple schemas and ensures that the most appropriate one—the one receiving the greatest amount of activation—is triggered. Contention scheduling ensures that you don't do two things at once, for example, trying to step on the accelerator and the brake while driving.

In the solving of problems that are related to new situations, existing schemas can fail, or there may be no learned schema that can be applied; that is, there is no a priori solution. In these cases, the contention system needs to be supplemented with an alternate selection process. According to the model, this supplementation occurs as the additional activation of other schemas from the SAS. The SAS contains general programming or planning systems that are slower and more flexible than the fast, automatic, and unchanging schemas of contention scheduling. These schemas are designed to be more general purpose and applicable to a wide variety of problem types.

So one can think of the SAS as a higher-level monitoring system, one that can apply general strategies to bear on a problem. The SAS can also suppress or turn off inappropriate schemas that might be triggered inadvertently. The failure to suppress these inappropriate schemas can result in "capture errors." Reason (1979) describes a person who, while walking across his back porch on the way to his car, began putting on his boots and jacket for working in the garden. Perhaps you can recall a similar situation from your own personal experience. Momentary lapses of attention seem to correspond to reduced activity or monitoring in the SAS.

So where in the brain are the different components of this model located? The left anterior frontal lobe is the probable location for the SAS. As mentioned above, lesions in this region are associated with difficulties in solving Tower of London problems. Damage to this region would explain the patterns of symptoms we see in executive dysfunction, where patients have trouble with problem solving but little difficulty in performing well-learned tasks. This would also explain environmental dependency syndrome. Here, the contention scheduling system is clearly operating, as environmental stimuli automatically trigger their learned responses. These responses are out of place, however, as they cannot be overridden by the supervisory system. It could also be that schemas or some aspects of schematic knowledge may be stored in the frontal lobes. This would follow, as disruptions to the frontal areas result in the failure to apply and use schemas and a consequent deficit in problem solving ability.

Stuss and Benson (1986) offer a second theory of executive function involving controlled and automatic attention. In their view, there is a three-tiered hierarchy of attentional systems. The automatic system that corresponds to the lowest level makes associations between sensory representations and other representations and is governed by posterior brain areas. The supervisory system that corresponds to the middle level runs executive processes and is used in problem solving. It resides in the frontal lobe. In addition, Stuss and Benson postulate a metacognitive system that corresponds to a third level. **Metacognition** refers to any process that monitors, regulates, or controls any aspect of cognition. Metacognitive regulation includes planning, resource allocation, checking, and error detection and correction (Brown, Bransford, Ferrara & Campione, 1983). The prefrontal cortex is believed to be the site of the metacognitive system. In a problem solving situation, metacognitive processing would evaluate whether a particular strategy is working and, if it decided that it was not working, would initiate consideration of another strategy. Individuals lacking metacognitive control would persist in their applications of inappropriate strategies; they would get stuck on one approach to a problem and fail to consider other options.

Overall Evaluation of the Neuroscience Approach

Part of neuroscience's job is a straightforward description of the human brain and the brains of other animal species that includes microscopic level descriptions of individual neurons and larger scale descriptions of brain structures. Although we have not mentioned it, comparative neuroscience compares these structures across species to see how evolutionary pressures have changed

them. In this sense, neuroscience and the evolutionary approach go very well together.

Physical structure of course cannot be divorced from function, and neuroscience also addresses physiology. But physiology is best described within some organizing framework. In recent years, that framework has become cognitive processes, and so we see the birth of a cognitive neuroscience that attempts to map cognitive function onto brain structure. Just as with the cognitive approach, we have seen, in the neuroscience approach, the creation of various models. These neuroscience models differ from their counterparts, described in the previous chapter, in that they specify the locations in the brain where information is represented and processed. They also specify the pathways by which information is passed from one processing center to another. The integration of the two types of model is a fruitful procedure for future investigations.

In Depth: Binding and Neural Synchrony

You may recall our previous discussion of the dorsal and ventral streams that process information concerning an object's location or motion, in the former case, and individual object features such as shape or color, in the latter case. If we were looking at a car going by in the street, the dorsal stream would contain representations of the car's location in space and its direction of motion. But the form and color of the car would be coded for by neurons in the IT cortex of the ventral pathway. This means that the information that represents an object is distributed across disparate areas of the brain. How does all this information come together to produce a unified perceptual object? This question of how to recombine the various features of an object is called **the binding problem**.

One solution has been posed by a group of German researchers (Engel et al., 1992; Singer, 1996). These researchers suggest that an object is represented by the joined and coordinated activity of a constellation of cells—a concept known as **neural synchrony**. Distinct cell subgroups of this constellation stand for individual features and may be separated by relatively large physical distances in the brain, but the dynamic activities of all of them represent an entire object. The same groups of neurons that code for a particular feature can participate in multiple constellations or cell assemblies and thus stand for the same feature as it is presented in other objects. For example, a network of cells representing the color red can participate in one assembly, for example, when one is looking at a red car. The same cells would then participate in another assembly, for example, when one is perceiving a tomato.

A problem with the existence of a group of functionally distinct cell populations in the brain is how they could stand out amid all the other ongoing activity. Singer (1999) proposes three solutions. First, other neurons not participating in the assembly can be inhibited. Second, the amplitude or strength of the cells in the assembly can be increased. Third, the cells in the assembly can synchronize their firing rates. This temporal synchrony means that all the participating neurons would fire at the same time. It is helpful to use an analogy here. Imagine a group of drummers in a band that has many musicians. If all the drummers banged out different rhythms they wouldn't stand out much. But, if the other musicians played more softly, the drummers played louder and, most importantly, began beating the same rhythm, we would then hear the drums as a salient ensemble.

There are some problems with temporal synchrony (von der Malsburg, 1981; Singer et al., 1997). Not all cells coding for features fire at the same rate. The processing speeds of these networks are also reduced. To overcome these difficulties, it was postulated that single or partial discharge rates of neurons, rather than their entire pattern of activity, can be synchronized (Gray, Konig, Engel, & Singer, 1989; Singer & Gray, 1995). In this scheme only one cell output or a small number of outputs that are part of a given sequence of outputs is timed to coincide with others. Returning to our musical example, it would be like having every fourth beat of one of the drummers coinciding with every fourth beat of the other drummers. Alternatively, two drummers could synchronize recurrent clusters of beats. This partial discharge synchrony would allow the drummers to play at different rates yet still coordinate their activity.

Although controversial, there is evidence to support the synchrony hypothesis. Engel et al. (1992) measured the electrical activity of spatially separated cells in the primary visual cortex. These cells respond individually to lines of a given orientation (Hubel, Wiesel & Stryker, 1978; Hubel, 1982). In this case both cells responded to vertical lines. When each cell was stimulated individually with a vertical line, the cells showed moderately correlated activity. But when a single prominent vertical line was used to stimulate both, they showed a much stronger coordinated activity. They fired short-lived bursts of output at the same time, alternating with periods of lowered activity. This implies that these two cells firing out of synchrony represent parts of an object, but firing synchronously are part of a larger assembly that codes for the entire object.

Minds On Exercise: Neural Functions

In this exercise, one group will work on attentional function. Another group will work on memory function. Additional groups may be formed for visual object

recognition and executive function—however, we have not surveyed all the areas that are involved in these processes and they would require additional research.

Each person in a group will represent one of the major structures involved in that function. In the case of attention, six individuals would represent the six major regions. All persons need to be able to answer the following questions: Where in the brain are they located? From what other brain structure(s) do they receive input? To what other brain structure(s) do they send outputs? What is the nature of these inputs and outputs; that is, what kind of information is it? Illustrate the information flow throughout the entire system using a concrete example.

Food for Thought: Discussion Questions

1. Why should faces be a class of visual object so important that a part of the brain seems devoted to perceiving them? Why is it important to be able to recognize and distinguish between faces? Think of the survival value of face processing for men, women, and children.
2. A picture of an apple is presented to the right visual field of a subject whose corpus callosum has been surgically cut. Would this patient be able to explain verbally what he or she has just seen? Could he or she identify an actual apple with one of his or her hands? Which hand? What if the image were displayed to the left visual field?
3. W. S., a 73-year-old man, has suffered a stroke. He can copy drawings of everyday objects successfully, but is unable to name these objects. What type of disorder does he suffer? How is his disorder different from others in its class?
4. You look at your cat "Scruffy" and reach over to pet him. This causes the cat to start purring. Where in your cortex is Scruffy's shape and color represented? His location and movement? The tactile sensation from stroking his fur? The motor command given to initiate the petting? The sound of the purring? How might all this information be brought together?
5. Based on the neurological evidence that is presented in this chapter, how many separate types of working memory do you think there are? What processes other than storage and rehearsal are associated with each?
6. In what ways (other than those already discussed) is the brain like a musical orchestra? In this analogy, what aspect of mind do the musicians stand for? The conductor? The music?

Go to the following website:
http://web.psych.ualberta.ca/~iwinship/studyguide/brain_study.htm

Run through the tutorial of the major brain structures, visible from a midsagittal section. What is the location of the hypothalamus in relation to the other structures? What functions does it perform? How does the hypothalamus maintain homeostasis? To what other brain area is it highly connected?

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Banich, M. T. (2004). *Cognitive neuroscience and neuropsychology*. Boston: Houghton Mifflin.

Benson, D. F., & Zaidel, E. (1985). *The dual brain: Hemispheric specialization in humans*. New York: Guilford Press.

Carter, R. (1998). *Mapping the mind*. Berkeley, CA: University of California Press.

Dudai, Y. (1989). *The neurobiology of memory: Concepts, findings, trends*. Oxford, UK: Oxford University Press.

Eichenbaum, H. (2002). *The cognitive neuroscience of memory*. New York: Oxford University Press.

Gazzaniga, M. S., & LeDoux, J. E. (1978). *The integrated mind*. New York: Plenum Press.

Goldberg, E. (2002). *The executive brain: Frontal lobes and the civilized mind*. New York: Oxford University Press.

Kandel, E. R., Schwartz, J. H., & Jessell, T. M. (1995). *Essentials of neural science and behavior*. New York: Simon & Schuster.

Kempler, D. (2004). Neurocognitive disorders in aging. Thousand Oaks, CA: Sage.

Posner, M. I. (2004). *Cognitive neuroscience of attention*. New York: Guilford Press.

Ramachandran, V. S., & Blakeslee, S. (1998). *Phantoms in the brain*. New York: Quill.

Sacks, O. (1985). *The man who mistook his wife for a hat*. New York: Summit Books.

7

The Network Approach: Mind as a Web

"The more the mind receives, the more does it expand."

—Seneca the Younger, 5 B.C.E.

The Network Perspective

The network approach is influenced by the principles of operation and organization of real world brains. All biological brains are made up of cells, called neurons, that are wired up to one another in a complicated fashion. Activity in one neuron or set of neurons in turn activates other neurons through these connections. It is this activity that underlies all mental operations—whether it be recognizing your cousin's face, calculating the tip on a restaurant bill, or deciding which law school to attend. **Connectionism** is a field of study in the network approach. Connectionists try to understand how the mind performs these kinds of operations via the construction of an **artificial neural network** (ANN), which is a computer simulation of how populations of actual neurons perform tasks. Semantic and propositional networks constitute another field of study under the network approach and are discussed later in this chapter.

The use of ANNs brings up a host of interesting issues in cognitive science. Perhaps the most interesting of these is the issue of knowledge representation.

The dominant paradigm in cognitive science is the use of symbols to represent information. These symbols are then represented, stored, and operated upon by cognitive or computer processes. In the case of ANNs, information is represented not in the form of a symbol, but as a pattern of activation in the network. The classical symbol representational view and the connectionist view differ also in their functional architecture. Traditional notions in cognitive psychology and machine intelligence have processing occurring in stages, whereby information is pulled from one relatively large-scale system and fed to another. In the network approach, by contrast, processing events occur in parallel and are mediated by many small processing units.

In the second part of this chapter we discuss knowledge representation in a new light. Most ANNs are capable of only limited representation—enough to enable the carrying out of tasks such as pattern recognition and classification. These limited representations do not adequately reflect the complexity and the considerable interrelatedness of the human conceptual faculty. To accommodate this limitation a different kind of network is needed, one that is capable of storing and using knowledge in the broadest sense. Semantic networks therefore model how information in a permanent memory store, such as human long-term memory, might be structured. They do this through their use of a rich set of interconnected concept and concept property nodes to represent information.

Principles Underlying Artificial Neural Networks

Traditional computers are **serial processors**. They perform one computation at a time. The result of a particular computing unit can then serve as the input to a second computation, whose new result serves as the starting point for yet another computation, and so on (see Figure 7.1). The brain as well as ANNs rely on a completely different processing strategy. Here, large numbers of computing units perform their calculations in parallel. One computing unit does not need to wait for another to finish its computation before it can begin its work. As shown in Figure 7.1, these units that operate in parallel are also not limited to receiving inputs from only a single unit: they can receive and process multiple inputs and transmit multiple outputs. This type of architecture is referred to as **parallel distributed processing**.

Researchers in the field of artificial intelligence and more generally in computer science solve a particular problem by constructing an algorithm, or procedure for solving it. Every detail and step of the procedure is planned or anticipated ahead of time. One conceptualizes the problem and its solution in

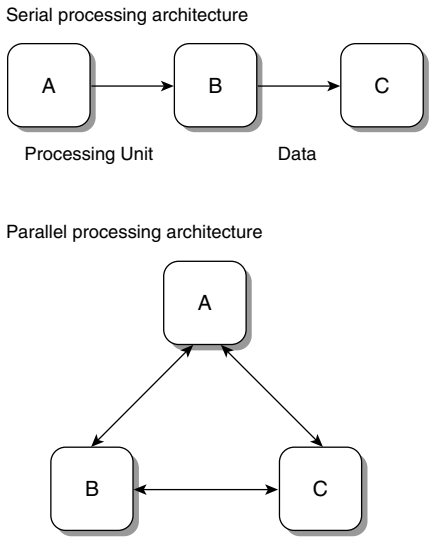


Figure 7.1 The top panel represents a serial processing architecture. Boxes are processing units and arrows are data. Each box performs a computation on the input it receives. The result of the computation is transmitted as an output to another processing unit. Unit B cannot start its computation until it receives its input from unit A. In the bottom panel is a parallel processing architecture. Each processing unit can be connected to every other. Computation in the units occurs simultaneously

terms of symbolic representations and transformations of these symbolic representations. This is referred to as the **knowledge-based** approach to problem solving, and it is used in many artificial intelligence approaches. In contrast, researchers using ANNs are more concerned with the overall behavior of a network. They leave the computational details up to the network itself and don't pay much attention to symbolic representations or rules. This is known as the **behavior-based** approach to problem-solving.

ANNs are pretty good at solving certain kinds of problems. Among them are problems of classification. These involve producing a learned label for a stimulus (pattern recognition) and assigning stimuli to categories (a form of concept formation). ANNs are also fairly adept at solving control problems, for example, the programming of the movements of a robot arm, and problems of constraint satisfaction, for example, the devising of airline schedules. In these situations, the data are often "noisy," the problem is usually ill-defined, and the outcome may be unclear. There are, though, many scenarios in which

the problem is well-defined and the solution process is clearly understood. In these scenarios, a conventional rule-based system like the ones used in artificial intelligence is preferred.

But let's get back to the issue of symbols. Why is it apparently O.K. for connectionists to ignore them? After all, it seems impossible that any kind of problem could be solved without concepts being represented in some fashion. The answer is that representations are inherent in ANNs, but do not exist in them in the form of symbols. They exist in most networks as a pattern of activation among the network's elements. This is known as a **distributed representation**. To illustrate, the idea of "banana" in a connectionist model might be represented by the simultaneous activity of the different elements or nodes A, B, and C and the connections among them. However, having said this, some ANNs do represent concepts via activity in single nodes. In these networks, activity in node A alone would represent, let's say, "apple." This is a form of **local representation**. In either case, though, we need not employ the idea of a symbol, such as a letter, to stand for the concept. The concept is instead represented as the activity of the nodes or node.

One advantage of neural nets is that they are capable of learning. That is, they can adaptively change their responses over time as they are presented with new information. However, it should be noted that ANNs are not the only form of simulated cognition capable of learning. Learning is also evidenced in machines that use symbolic methods. The acquisition of new information or a new skill usually requires repetition. For example, a child learning the multiplication tables will repeat them over and over to herself. If she makes a mistake, she must correct herself and then repeat the new information. Neural networks operate on these same principles. They learn over a series of trials to perform a task or to come up with an answer to a question. If they produce a wrong answer, the correct answer can be "shown" to them. This feedback is then used to adjust the performance of the network until it produces the correct answer.

Characteristics of Artificial Neural Networks

Real neural networks exist in the brain in the form of neurons and the connections between them. The artificial networks constructed by connectionists exist only as software simulations that are run on a computer. Each neuron, or basic computing unit in an artificial network, is represented as a **node** and the connections between nodes are represented as **links**. A node, if it is stimulated,

sends out a signal, represented as an activation value, which runs along the link that connects it to another node or nodes. A node follows a set of internal rules that "decide" if it is to fire. The simplest of these rules is: the node fires if the input it receives exceeds a **threshold** value. If the input is greater than or equal to the threshold, it fires. If the input is less than the threshold, it does not.

Links in a neural network have **weights**, which specify the strength of the link. A weight can be positive, negative, or zero. The numeric value of a weight runs between 0 and 1.0: the higher the numeric value, the heavier the weight. The net output of a unit is its activation value multiplied by the weight of the relevant link. So, for example, a unit with an activation value of 1 passing along a link that has a weight of 0.5 will have a net output of 0.5. It will positively stimulate the node to which it is connected by a factor of 0.5. The greater the value of a node's net output in the positive direction, the more likely it is that the nodes it is connected to will fire.

A unit with an activation value of 1 passing along a link that has a weight of –0.5 would have a net output of –0.5, which would negatively stimulate whatever nodes it feeds to. The greater the value of a node's net output in the negative direction, the less likely it is that the nodes it is connected to will fire. Negative output thus serves the function of dampening or shutting down the activity of other nodes. This is similar to the role of inhibition in biological networks, whereby one neuron can "turn off" or slow down another. The output of any node that is part of a link that has a zero weighting is of course zero, meaning that there is no effect on any downstream nodes. If a node receives two or more outputs from other nodes, it takes these outputs and, in effect, summates them to determine whether it should fire. Figure 7.2 shows a simple neural network with activation values, weights, and net outputs.

The amount of stimulation a given node receives is specified by a **basis function**. It receives the input signal, which can be in the form of an input to the network or the output of another node. The basis function calculates a simple summation of all the inputs the node receives. This can be described by the following equation:

$$S_j = \sum_{i=0}^{n} a_i w_{ji}$$

where w_{ji} = the weight associated with the connection between processing unit j and processing unit i, and a_i is the value of the input signal i. The results are summed over all inputs to the given node.

The basis function next sends its input signal to an **activation function**. This function maps the strength of the inputs a node receives onto the node's

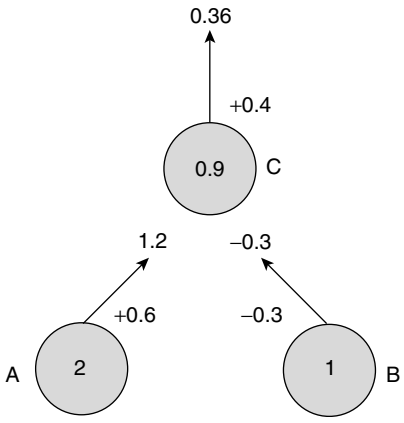


Figure 7.2 A simple neural network. Activation values are shown inside each node. Weights are indicated on the links and net outputs at the endpoint of each connection. Node A has an activation value of 2 and a weight of +0.6. Its net output is thus 1.2. Node B's activation value is 1 and its weight is –0.3, yielding a net output of –0.3. These outputs are summated at node C so that it has an activation value of +0.9. The activation value of node C multiplied by its weight of 0.4 produces an output of +0.36

output. Most activation functions are sigmoidal (S-shaped) and are specified as:

$$f(x) = \frac{1}{(1 + e^{-x})}$$

where x is the input value. Figure 7.3 shows the shape of the sigmoid activation function.

Early Conceptions of Neural Networks

Warren McCulloch and Walter Pitts were the first researchers to propose how biological networks might function, in 1943. They made a few simple assumptions about how neurons might operate. They assumed that each neuron had a binary output, that is, it could either send out a signal, corresponding to its being "on," or not send out a signal, corresponding to its being in an "off" state. Whether or not a neuron would fire was determined by a threshold value. The weights of the connections between neurons, in their model, were

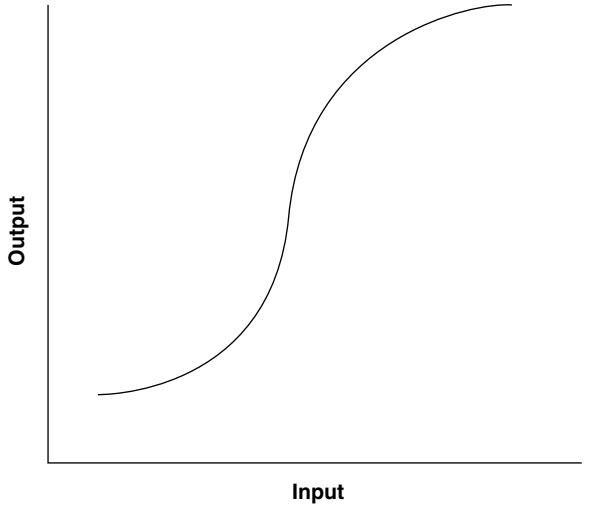


Figure 7.3 A sigmoid activation function. As the input to a node increases, the likelihood that the node will fire also increases

additionally assumed to be at a fixed value. Networks operating under these rules are capable of computing simple logical operations, such as OR, AND, and NOT. A neuron programmed to compute a NOT operation will fire if it receives activation at or above threshold. It won't fire if it receives inhibitory input. Digital computers of the sort that sit on our desks perform calculations that are based on these simple logical functions. This means that a neural network capable of performing these calculations can in theory do anything a digital computer can.

Donald O. Hebb (1949) was the first person to propose how changes among neurons might explain learning. According to the Hebb rule, when one cell repeatedly activates another, the strength of the connection between the two cells is increased. In this fashion, pathways or circuits among neurons are formed. These circuits are believed to be the neural foundation of learning and memory. Imagine being at a party and trying to remember someone's name. How would you do it? Imagine at the party you are trying to remember a phone number you had been given. If you could not write the number down, you would repeat it over and over—a process called rehearsal. Each repetition would correspond to a new activation of the circuit, strengthening it further. The circuit itself, once strong enough, would represent the telephone number. Retrieval of the number at a later date would then be equivalent to a reactivation of the circuit.

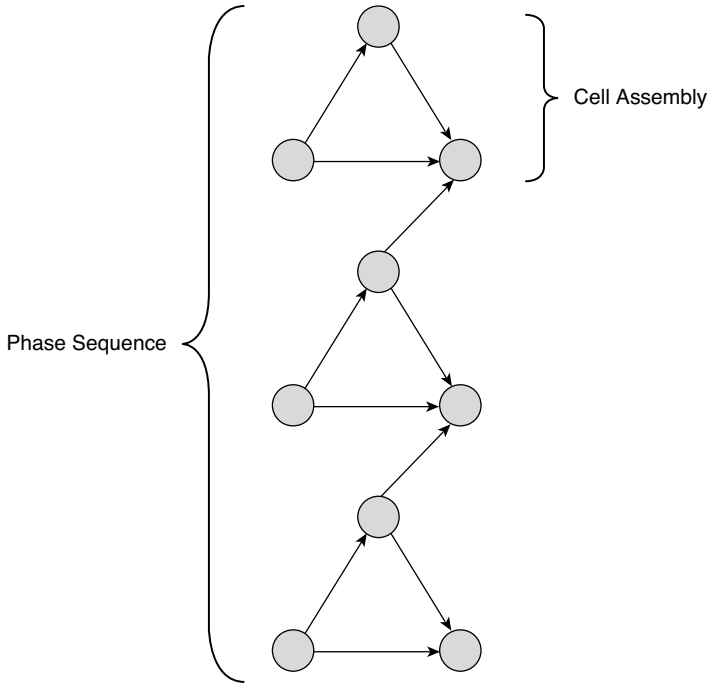


Figure 7.4 A phase sequence consisting of cell assemblies

Hebb defined two types of cell groupings. A **cell assembly** is a small group of neurons that repeatedly stimulate each other. A **phase sequence** is a group of connected cell assemblies that fire synchronously or nearly synchronously (see Figure 7.4). If a cell assembly coded for a simple perceptual quality, such as "red" or "round," then these qualities could become linked so as to form a phase sequence during learning and code for a higher order concept such as "apple."

Beginning in the 1950s, research on neural nets focused less on logical operations and more on mimicking real biological function. An artificial nervous system called the **perceptron** was introduced (Rosenblatt, 1958). Perceptrons are neural nets designed to detect and recognize patterned information about the world, store this information, and use it in some fashion. Perceptrons are also characterized by their ability to learn from experience: they can modify their connection strengths by comparing their actual output to a desired output called **the teacher**. The ANN networks discussed in subsequent parts of this chapter constitute different types of perceptrons.

The earliest perceptron was an artificial retina called the "Mark I" (Rosenblatt, 1958). This network could recognize simple visual patterns, such

as vertical and horizontal lines. It was capable, with training, of producing a desired output for each different type of visual pattern. The first perceptrons were quite simple. Each contained a single layer of input units or an input and an output layer (see Figure 7.5). The limitations of these fledgling perceptrons soon become clear. A major flaw was their inability to distinguish among certain patterns (Minsky & Papert, 1969). This was in part due to their relatively weak computing power. A single layer or two layers and the connections between them do not provide for much in the way of complexity and flexibility. The solution, of course, was to build more complicated networks.

Back Propagation and Convergent Dynamics

In a three-layer network, the computing units or nodes are organized into three distinct groups. A representation of the stimulus is presented to the **input layer**. These units send signals to a **hidden layer**, which in turn feeds activation energy to an **output layer**. The output layer generates a representation of the response. Figure 7.6 depicts a three-layer network. For a detailed description of a three-layer network that learns through back propagation, read about NETtalk in the In Depth section.

Here's how a three-layer network designed to recognize letters works: The capital letter "A" activates the nodes in the input layer. These nodes send activation energy via links to the hidden layer. The nodes in the hidden layer send signals via their links to nodes in the output layer. The pattern of activation in the output layer is the network's initial response to the letter. This response is then compared to the target response, represented by the teacher. The difference between the actual and desired outputs, the **error signal**, then feeds back to the output layer. The network uses the error signal to modify the weights of the links. Figure 7.7 shows these different steps that are part of the training of a three-layer perceptron. The modified weights allow the network (the next time it "sees" the letter "A") to generate a response that is closer to the desired one. After repeated presentations of the stimulus in the presence of feedback, the network is able to produce the target response. It has, in effect, learned to recognize the letter. This kind of training based on error feedback is called the **generalized delta rule** or the **back-propagation** learning model.

Using the simplest perceptron learning rule, a constant is added to or subtracted from the appropriate weights during learning:

$$w_{ji}^{new} = w_{ji}^{old} + C(t_j - x_j)a_i$$

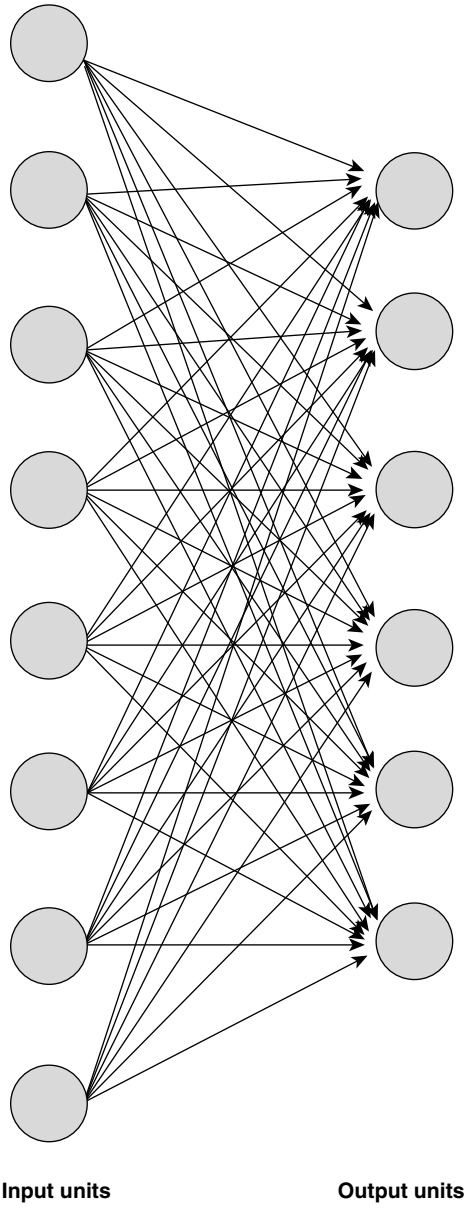


Figure 7.5 An example of an early perceptron with two layers. Notice each input unit maps onto every output unit

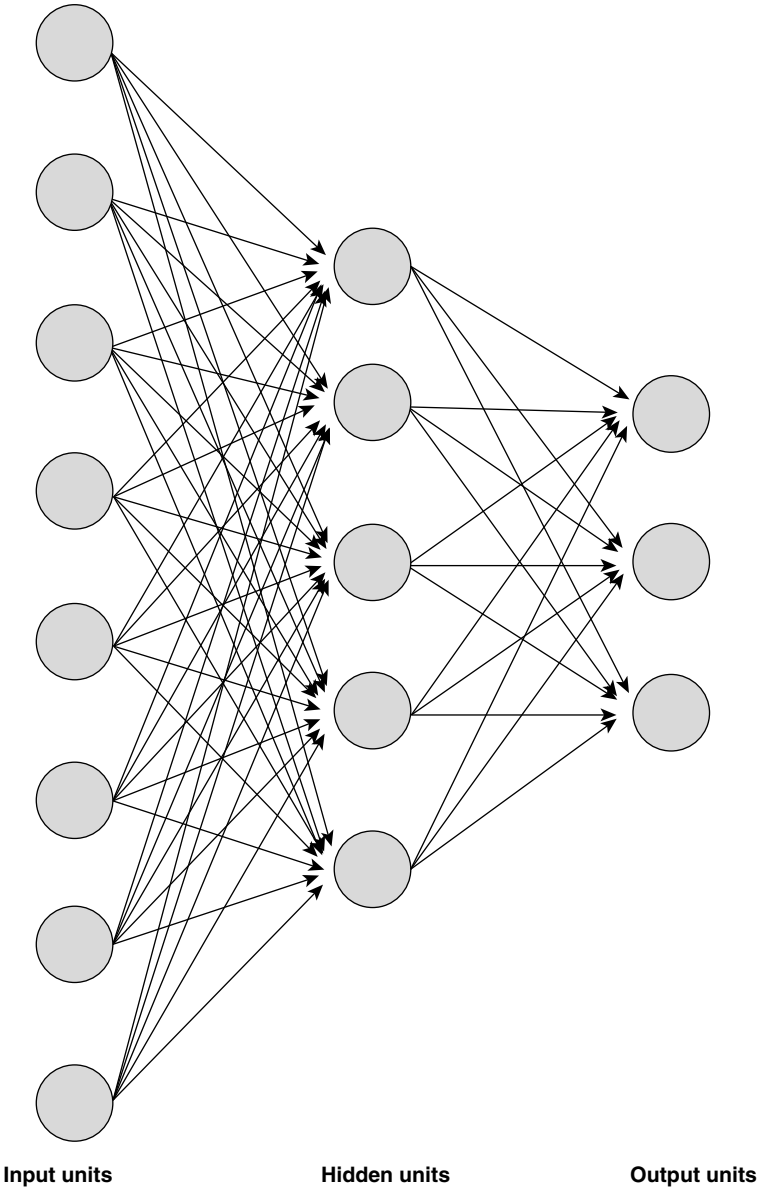


Figure 7.6 A three-layered neural network with input, hidden, and output layers

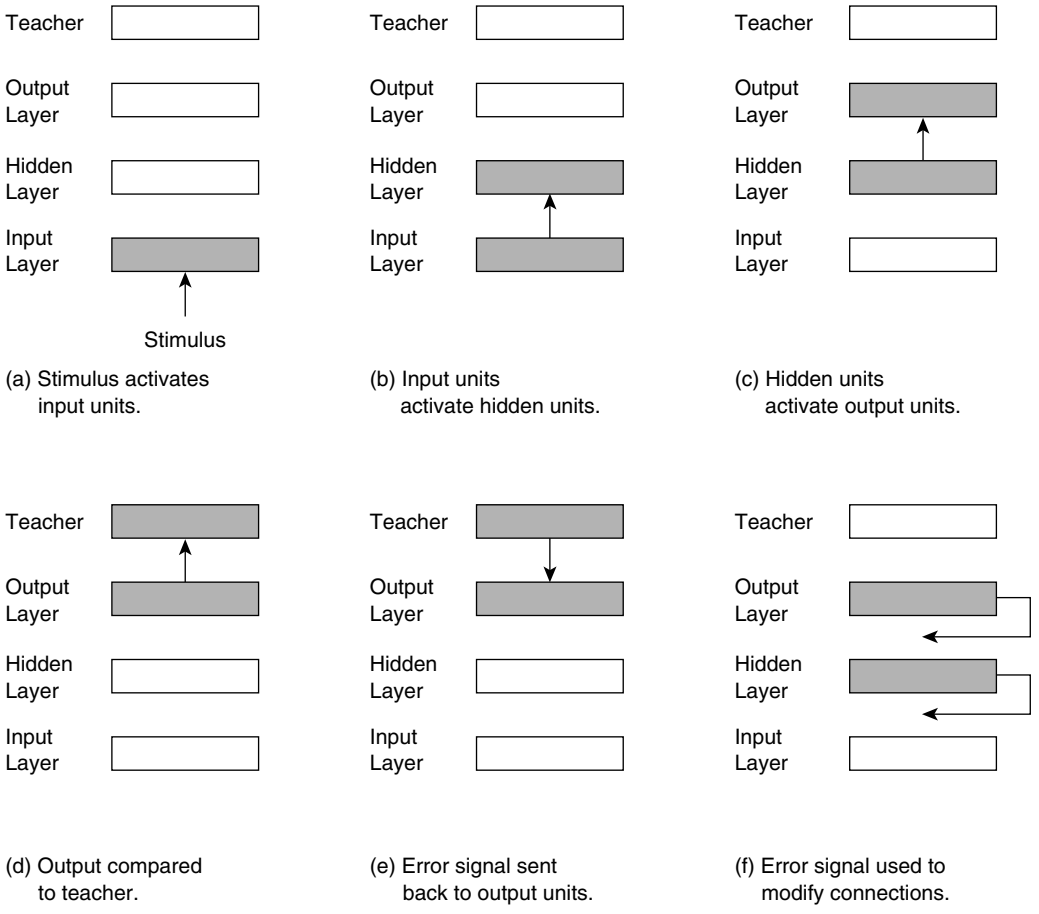

Teacher
Output Layer
Hidden Layer
Input Layer
Stimulus
(a) Stimulus activates input units.
(b) Input units activate hidden units.
(c) Hidden units activate output units.
(d) Output compared to teacher.
(e) Error signal sent back to output units.
(f) Error signal used to modify connections.

Figure 7.7 Steps in the training of a three-layered network using the back-propagation learning model

The term C is a constant, usually set at below 1. It is equivalent to the "learning rate." The greater the value of C, the greater the size of the adjustment made to the connection weight on any given round of computation. The term t is the target value for output unit j following presentation of the pattern, and x is the actual value of the output of output unit j following presentation of the pattern. There must be a difference between these two terms for any error correction to occur. This formula also requires that the input unit be active before a weight is changed. The term a represents the activity of the input unit and is either 1 or 0. If there is no activity, a's value is 0 and no change is made to the weight.

One rough way of considering ANNs is according to their **convergent dynamics** (Braspenning, 1995). This refers to how the state of the network—essentially the values of its weights—changes over time. It turns out that networks display three basic types of dynamic:

1. *Convergent*. Most ANNs exhibit convergent properties. There is a significant amount of activity in the network at first. This activity then slows down until the network settles into a stable state.
2. *Oscillatory*. Here, the weights fluctuate periodically—they both increase and decrease—in a fairly regular fashion over time.
3. *Chaotic*. In this dynamic, the network's activity varies in a chaotic fashion; it shows periodic and non-periodic fluctuations.

A deficiency of some ANNs concerns this dynamic behavior. In some networks, the activity level settles down too soon. One way of assessing this is to plot a **loss function**—the change in the error signal over the set of learning trials. In some cases, we might then see that the error level drops very quickly. If this occurs, the network is not able to perform its task and is caught in what is called a **local minimum**. There are various ways to correct for local minima. They include restarting the learning process at a different point and using noise to "shake" the network state out of its problem state.

Artificial Neural Network Typologies

Over the past several decades, a plethora of different types of networks have been developed. A variety of organizing schemes are used to classify these different types. Hudson and Potsma (1995), Maren et al. (1990), and Dayhoff (1990) all use different ordering criteria. To simplify the matter, we first list the criteria or major characteristics of the various types of networks and then describe prominent examples of these types. Each of the examples is described here in summary fashion. The interested reader is asked to consult the source material to learn more.

1. *Supervision*. Networks can be put into two broad categories, based on the way they learn. **Supervised** networks are presented with target answers for each pattern they are given as input. That is, the network "knows" what the right answer is in each training trial. In **unsupervised** networks, the network must determine the answer on its own, without the benefit of an answer.

2. *Layers*. Networks vary in the number of layers they contain. Broadly, they can be categorized as either **single-layer**, in which case they have only one layer of nodes, or **multi-layer**, in which case they have two, three, or more layers.

3. *Information Flow*. Information in a network is either feed-forward or recurrent. In a **feed-forward network**, the flow of activation is in one direction only—forward; flow is from units in an input layer to units in other layers. Information in feedback or **recurrent networks** can flow in two directions, both forward and backward. The backward flow, for example, can be from output layers to hidden or input layers.

Perceptrons have already been described. They are supervised networks, as the teacher provides the correct response and allows for error adjustment of the weights during learning. They are multi-layer: they contain either two or three layers of nodes. In back-propagation networks, even though the results of forward signal passing may be used to correct the weights of the links of previous nodes, the operations of the net are strictly feed-forward. Perceptrons are the typical network of choice for the task of pattern classification. Three-layer perceptrons acquire the ability to recognize the more general features of their input patterns. Their weakness is that they often require long training periods.

Hopfield-Tank networks are a type of supervised, single-layer, laterally connected networks (Hopfield and Tank, 1985). Because there is only one layer, these nodes, by default, can only be connected to one another. Most are fully connected, meaning that each node is connected to every other node. These networks are autoassociative—they are good at regenerating clean versions of patterns they have prior experience of when presented with noisy or incomplete versions of those patterns as input. Hopfield-Tank nets are also good at solving optimization problems, which involve coming up with the right solution in the presence of a large number of possible combinations. In their unmodified state, they have trouble avoiding local minima.

We now turn to unsupervised networks. A **Kohonen network** is an example of a two-layer network (Kohonen, 1990). These nets are also called feature maps. They are able to create a topological map or spatial representation of the features that are present in the stimulus input. These maps or representations are quite similar to the topological maps that exist in the brain. The primary visual cortex, for instance, represents various characteristics of visual input, including ocular dominance columns, which process information from each eye, and orientation columns, which represent different line orientations. The actual architectures that a Kohonen network and the brain use to accomplish this, however, are different.

The **Adaptive Resonance Theory network** (ART) is an example of an unsupervised, multi-layer, recurrent network (Carpenter and Grossberg, 1988). It is able to classify input patterns and put them into different categories in the absence of a teacher. It does this by employing a resonant dynamic, in which the patterns in the first and second layers stimulate each other repeatedly, until the pattern in each layer settles into a stable state. A disadvantage of the ART network is that individual nodes represent categories. This local representation makes the net particularly susceptible to degradation with damage, i.e., when the nodes representing categories are destroyed, information about the categories is lost.

Evaluating the Connectionist Approach

Advantages

The connectionist approach has many strengths. The most important one is the similarity between network models and real live neural networks—what is termed **biological plausibility** (Stillings, Weisler, Chase, Feinstein, Garfield & Rissland, 1995). This plausibility can be demonstrated in three fundamental ways. First, artificial networks share general structural and functional correlates with biological networks. Second, and as pointed out previously, artificial networks, like their biological cousins, are capable of learning. Third, artificial networks react to damage in the same way that human brains do. In this section, these advantages and other advantages of networks are discussed in depth.

Many of the elements of artificial networks have their counterparts in biological networks. The nodes are equivalent to neurons, whereas the links are analogous to axonal and dendritic connections. The functioning of constructed neural networks is also very biological. At a microscopic scale, the decision rules that specify when and how an individual node fires are based on neuron physiology. At a macroscopic scale, the parallel, distributed processing of connectionist networks is also found in the brain.

Learning in connectionist models takes place via the adjustment of the weights of the links between nodes. Work in neurophysiology shows that this process, at least in simulation, is not too far off from what happens in real brains. Recall from the preceding chapter our discussion of synaptic plasticity and of how it is mediated by long-term potentiation. If two cells in a circuit are simultaneously active, the synapse between them is strengthened. This increase in synaptic strength corresponds to a numerical increase in the weight of a connection between nodes in an ANN.

A third argument in support of the biological plausibility of connectionist models is the way they respond to damage. Neural networks demonstrate **graceful degradation**. This is a gradual decrease in performance with increased damage to the network. In graceful degradation, small amounts of damage engender only small reductions in performance, whereas greater damage produces correspondingly larger deficits. Human patients who have sustained brain damage show evidence of the same phenomenon. Graceful degradation is a property of a parallel computing architecture. If some nodes are destroyed, others can "take over." In other words, the network can adapt to the damage by calling on its distributed function across the existing nodes and links. Of course, this advantage holds only for nets with distributed representations. Those networks with localized representations, such as the ART net, are more subject to failure in the event of damage.

Connectionist networks display two other interesting "psychological" phenomena: interference and generalization (Green & Vervaeke, 1996). **Interference** refers to instances in which two sets of information that are similar in content interfere with one another. If a student were studying Spanish and Italian at the same time, he might forget words from one language that are similar in sound to words in the other. Small networks trained to learn large numbers of patterns show signs of interference: they have difficulty in distinguishing similar patterns. **Generalization** is represented by the ability to apply a learned rule to a novel situation. If one assumed that any animal that had wings was a bird and then saw an animal for the first time that also had wings and called it a bird, he or she would be generalizing. Rumelhart and McClelland (1986) trained a network to learn the past tenses of verbs. The network learned to conjugate new verbs. It was able to produce "wept" for "weep" because it had been trained to associate "slept" with "sleep." The network was capable of bona fide generalization.

Problems and Disadvantages

We have looked at the many advantages of the connectionist approach, but it is also limited, in a number of respects. What at first seems like its primary strength, biological plausibility, should also be viewed as problematic. If we make a more detailed comparison of ANNs and actual brains, we discover a number of significant differences. First, real neurons are massively parallel—they exchange signals with thousands of others. It is not yet possible to simulate parallel processing of this magnitude; such processing would require simply too many nodes and connections and too much processing power. It is

likely that some of the emergent characteristics of the brain arise from massive parallelism. Until we are able to build ANNs of this complexity, we may not be able to understand these emergent characteristics.

Second, most networks, as stated earlier, show a convergent dynamic. The activity of such a network eventually dies down and reaches a stable state. This is not the case for brain activity. Real neural networks are oscillatory and chaotic—their states fluctuate over time and do not settle. Neural activity in the hypothalamus, for example, is oscillatory. As regards this type of dynamic, it is a fascinating but difficult question to determine where in the network a representation exists. In convergent networks, the presentation of a stimulus pattern reliably elicits the trained response. This response is the network's distributed representation of the correct answer in the form of a pattern of activation among nodes. But in oscillatory and chaotic networks, there is no such representation. The network, because it is constantly changing, has not formed any specific pattern of activation that can be linked to a stimulus. Representations in these systems might correspond to more global characteristics, such as the frequency or phase of the network's activity.

Dawson (1998) points out an additional problem. He discusses Elman's recurrent-type network (1990), which has difficulty in analyzing inputs that are staggered in time. Dawson speculates that the problem lies either in the architecture of the network itself or in the learning rule. In the case of Elman's network, it turned out to be the latter, because in instances in which a new rule was implemented, the network was adequate to the task (Kremer, 1995). We see here another difficulty, which is that networks may have inadequate learning rules. The error correction procedures for adjusting weights that are currently in use represent only one of many possible ways of training a network. A subject for future research would be the exploration of alternate learning methodologies.

One specific problem that arises in network learning is the **stability-plasticity dilemma**. It states that a network should be plastic enough to store novel input patterns; at the same time it should be stable enough to prevent previously encoded patterns from being erased (Grossberg, 1987). This apparent conflict is analogous to the phenomenon of psychological interference, discussed previously. The fact that ANNs show evidence of being caught in this dilemma is useful, because it may offer some insights into human interference. But it becomes a real problem when one is attempting to implement artificial networks. **Catastrophic interference** occurs in instances in which a network has learned to recognize a set of patterns and then is called upon to learn a new set (French, 2001). The learning of the new set modifies the weights of the network in such a way that the original set is forgotten. In other words, the newly

learned patterns suddenly and completely ("catastrophically") erase the network's memory of the original patterns. A variety of solutions to the problem have been proposed, ranging from the introduction of "novelty vectors" to a reduction in the internal overlap of representations, to pre-training the network with random samples of patterns (French, 1992; Kortge, 1990; McRae & Hetherington, 1993).

We should discuss learning a bit further. In supervised networks, a "teacher" or training pattern is necessary in order for the network to learn. But where does this teacher come from? Humans and other animals learn in many instances in which there is no right answer provided, or in which the answer is less accurately provided (Barto, Sutton, & Anderson, 1983). Also, there is no evidence that biological networks feed an error signal back to "previous" units to modify connection strengths, as in the back-propagation model (Bechtel & Abrahamsen, 1991). This is simply a convenient engineering solution to the problem, but fails to address how an analogous process in actual brains might occur.

Semantic Networks: Meaning in the Web

There is another class of network models that have many features in common with neural nets. Nodes, links, thresholds, and summation of input also characterize these networks. They are **semantic networks**. But in semantic networks, each node has a specific meaning. Semantic networks therefore employ the local representation of concepts. Semantic networks have been adopted primarily by cognitive psychologists as a way to explain the organization and retrieval of information in long-term memory. In this section, we will discuss the characteristics of such networks, drawing on experimental results from cognitive psychology.

One motivation for studying semantic networks centers on representational capacity. The distributed representations formed in convergent ANNs are very simple and fail to capture the richness of human conceptual capacity. Even if it were to turn out that humans store concepts as distributed patterns in a neural circuit, it would be unclear as to how this kind of storage might account for the complexity of our concepts. When we think of a concept such as "dog," we are doing much more than coming up with a label. Our concept of "dog" goes beyond pattern recognition to include semantic content—what dogs look like, what they do, what our own experience of dogs is, and so on, and is intimately tied up with other concepts we possess. Semantic networks allow us to represent and understand these more complex aspects of concepts.

Characteristics of Semantic Networks

In semantic network models, a node's activity can spread outward along links to activate other nodes. These nodes can then activate still others—a process called **spreading activation**. An assumption of some models is that activation energy decreases with increasing distance, since it is believed that the activation energy encounters resistance as it passes through succeeding links and nodes. This means that spreading activation may lose strength as it travels outward from its point of origin. Another characteristic of these networks is that the distance between two nodes is determined by their degree of relatedness. Concepts such as "automobile" and "truck" are semantically related and hence fairly close to each other, whereas concepts such as "automobile" and "flower" are less related and are farther apart. It is therefore easier to activate the "truck" node via the "automobile" node than it is to activate the "flower" node in this way.

Spreading activation is thought to underlie retrieval of information from long-term memory. For example, suppose a test question in history class asks for the names of major western European nations. While you were studying, you had memorized these nations in several ways: on the basis of their geographic locations, whether they fought each other in wars, how related their languages are, and so on. You start by naming Germany, then Italy. At the end, you realize you've left out a country, but can't think of what it is. Finally, it comes to you. You remember that the country is France—because while you were studying, you realized that France and Germany fought each other during the two world wars. Although you were not able to retrieve France on the basis of the Mediterranean characteristics it shares with Italy, you were able to retrieve it on the basis of its historical relation with Germany. This underscores an important principle of effective studying: it is best to associate a concept with as many other related concepts as possible, to ensure a higher likelihood of retrieval. In this way, there exists a greater number of alternative pathways that lead to a given node. If one nodal connection fails to activate the target item, spreading activation will ensure that some other pathway does. These alternate associations that facilitate recall are also called **retrieval cues**. Figure 7.8 illustrates how a semantic network might represent knowledge of western European countries.

An important characteristic of semantic networks is **priming**. In priming, the processing of a stimulus is facilitated by the network's prior exposure to a related stimulus. A study by Meyer and Schvaneveldt (1971) demonstrates this phenomenon. The participants in their study were presented with pairs of letter strings. Each letter string was either an actual word or a nonsense word.

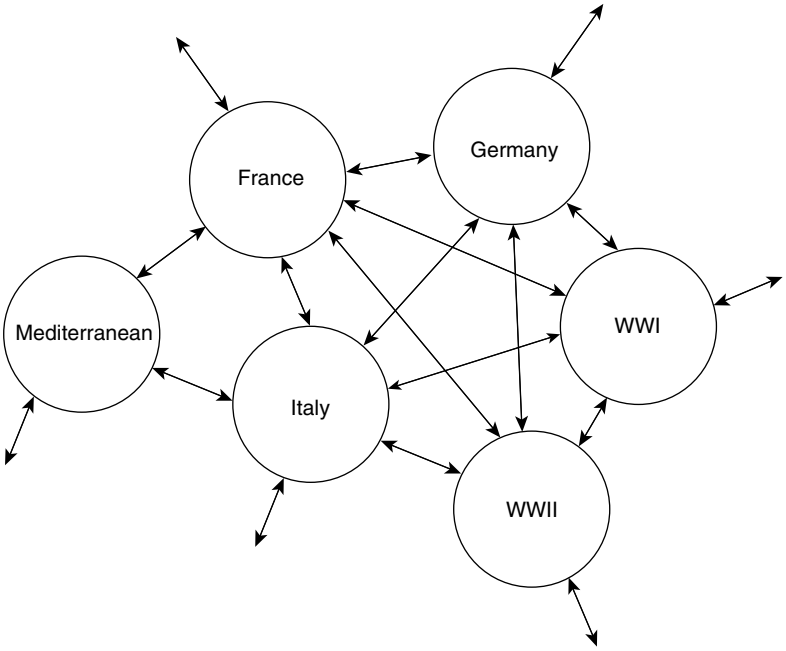


Figure 7.8 A hypothetical example of how knowledge of western European countries might be represented in a semantic network

Participants were asked to push a button when they encountered actual words. The button push recorded how long it took them to respond. If both words in a pair were meaningful and related, the participants were faster to respond correctly to the second word in the pair. So, if the first word was "Nurse" and the second "Doctor," responses were faster than if the first word was "Nurse" and the second "Butter." The explanation goes as follows. The "Nurse" node in long-term memory becomes activated first. Through spreading activation, it activates other, semantically related nodes, such as "Doctor." A short time later, when the participant's response to "Doctor" is required, the response time is reduced because that node is already partially activated. This technique can be used to map out the structures of semantic networks. The larger the priming effect, the closer together the relevant nodes should be (McNamara, 1992).

A Hierarchical Semantic Network

A classic study by Collins and Quillian in 1969 suggests that semantic networks may have a **hierarchical organization**, with different levels representing

concepts ranging from the most abstract down to the most concrete. They used a **sentence verification** task. Participants were asked to respond to individual sentences that appeared on a computer screen. If they judged a sentence to be true, they hit one button. If they judged it to be false, they hit another. Examples of true sentences were "A canary is a bird" and "A canary is an animal." Response times were recorded.

Collins and Quillian theorized that a correct response to one of these sentences required overlap of spreading activation. To know whether a canary is a bird would require the activation of both a "canary" node and a "bird" node. Recognition of both concepts would activate these nodes. Activation would then radiate outward through the network from each node until each individual unit's activation would mutually affect one another. When the activation of these two nodes overlaps, the participant knows that the two are related and can then confirm the sentence. If the nodes have a close semantic relation, they should be in proximity to each other in the network and responses will be fast because spreading activation will have less distance to cover. If the nodes are less related, the distance between them will be greater and response times will be longer.

Based on the set of reaction times they obtained in the study, Collins and Quillian sketched out a hypothetical memory structure for knowledge of animals. This is shown in Figure 7.9. The hierarchical nature of the network is immediately clear. The concept "animals" and its properties, such as "eat food," "breath," and "have skin," are at the top of the hierarchy. Because the idea of animals is abstract and encompasses all known types of animal, "animals" constitutes a **superordinate** category. At the next lower level we have examples of classes of animals, such as "birds," "cats," and "dogs," accompanied by nodes corresponding to their characteristics. For birds, these would be "can fly" and "lays eggs." Because these different classes of animal are at a moderate level of specificity, they are **ordinate** categories. At the bottom of the hierarchy are nodes that are even more concrete, corresponding to exact species of animals; these animal species are **subordinate** categories. Properties of these animal species are also at this same level. A node corresponding to "canary" would have links at the subordinate level connecting it to "can sing" and "is yellow."

Figure 7.10 depicts a graph of the reaction times that led to the model's being proposed. Responses for which spreading activation occurs entirely within a level were quite fast. It took participants about 1,000 milliseconds, on average, to confirm that "A canary is a canary." In this instance no effective spreading activation is required, the statement is self-referential, and only the canary node is active. To confirm "A canary is a bird" took slightly longer—about 1,200 milliseconds. Presumably, the longer time is owing to the fact that both the canary and bird nodes must be activated and spreading activation must travel

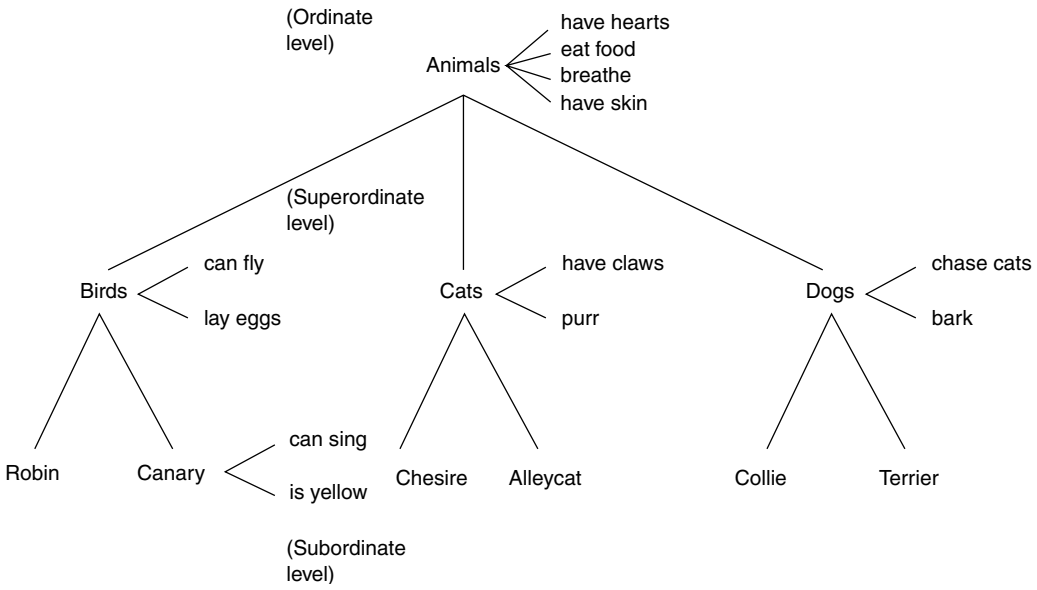


Figure 7.9 The hierarchical semantic network proposed by Collins and Qullian. Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory

Source: Collins, A.M., & Quillan, M.R., Retrieval time from semantic memory in *Journal of Verbal Learning and Verbal Behavior, 8*, pp. 240-247, copyright © 1969, with permission from Elsevier.

along a link, between the subordinate and ordinate levels. Verification of the statement "A canary is an animal" took even longer. In this instance, the canary and animal nodes are activated and spreading activation must travel along two links, from the subordinate level to the superordinate level.

Finally, you will notice that response times for property statements follow this same function. But the verification of these statements is apt to take even longer. Sentences like "A canary can sing" were quickly verified, because the "can sing" node that defines canary properties is at the same level as the "canary" node. Sentences like "A canary can fly" were verified less quickly, because the property node "can fly" is attached to the "bird" node at the ordinate level. In general property statements take more time because activation must spread from property nodes to concept nodes—within a level as well as between concept nodes that occupy different levels.

Evaluating the Hierarchical Model

The hierarchical model is intuitively appealing but fails to account for several findings (McCloskey & Glucksberg, 1978). Participants are faster to

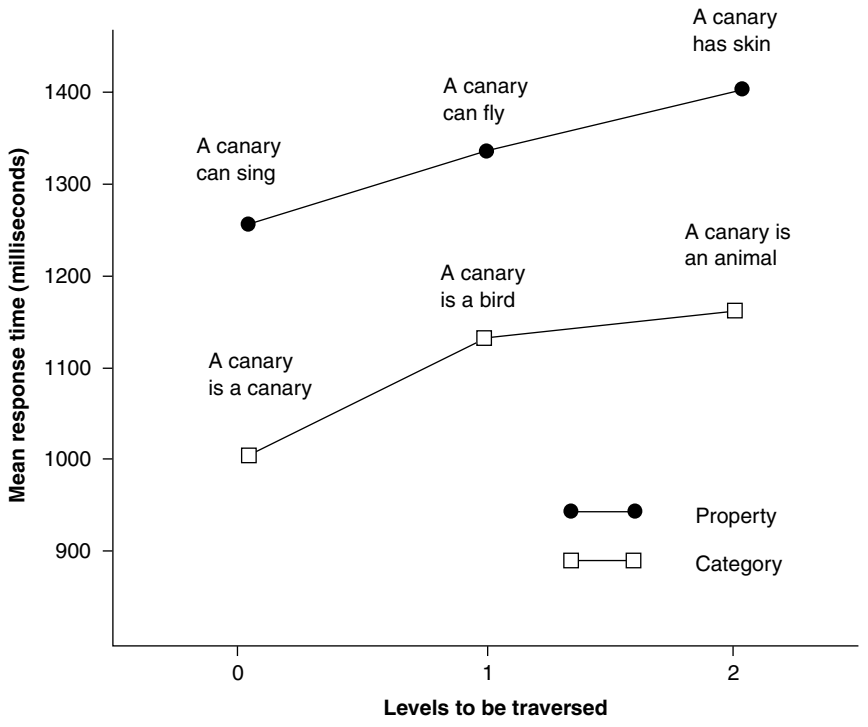


Figure 7.10 A graph depicting the results of the Collins and Quillian experiment. Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory

Source: Collins, A.M., & Quillan, M.R., Retrieval time from semantic memory in *Journal of Verbal Learning and Verbal Behavior, 8*, pp. 240-247, copyright © 1969, with permission from Elsevier.

make the correct response to "A dog is an animal" than to "A dog is a mammal," even though "animal" is farther away from "dog" in a hierarchical scheme. Evidently, people tend to think of dogs more as animals than as mammals, even though "mammals" is a subset of animals and should have greater proximity to the concept of dog than does "animals." Also, the sentence "A robin is a bird" is responded to more quickly than "An ostrich is a bird," even though the two sentences are equivalent and ostriches and robins are both subordinate to the bird category. This suggests that some birds are more "typical" than others, and that concepts may be represented by **prototypes** that represent generic or idealized versions of those concepts. If this were the case, then the network would need restructuring and the "robin" node would have to be moved—so that it was closer to "bird" than was the "ostrich" node. These results imply that the hierarchical organization of the sort seen in the Collins

and Quillian model may be overly contrived and may represent only one of the ways in which we organize concepts.

A final criticism of the hierarchical model has to do with the principle of **cognitive economy**. This principle states that nodes should not have to be coded for more times than is necessary. Collins and Quillian attempted to preserve cognitive economy by placing property nodes only at their appropriate levels in the hierarchy. For instance, "can fly" is positioned as a property of birds only, even thought it could be linked to the canary node and to nodes representing every other kind of bird as well. The advantage of cognitive economy is obvious. It eliminates redundancy and frees up resources in a memory system. But research has found no difference in the response times for the sentences "A bird has feathers" and "A canary has feathers." So the property "has feathers" seems to be attached to our concept of canary as well as that of birds. Like the network hierarchy, cognitive economy seems to work better in principle than in reality.

Propositional Semantic Networks

Semantic networks of the sort described above are fine for representing simple factual properties of objects in the world. They can represent a category relationship, which is depicted by an "**isa**" **link**. The connection between the "bird" and "animal" nodes would be this type of link, as "A bird *isa* animal." They can also represent a property type relationship with a "**hasa**" **link**. The "bird" and "feathers" nodes would be connected by a hasa link, since "A bird *hasa* feathers." But how would we set up a network to stand for more complicated relationships among concepts? What would the network that could encompass the sentence "The dog chased the cat" look like?

The answer is a new class of network, designed to code for propositions. John Anderson has developed such a network as part of his ACT* model of cognition, which was discussed in the cognitive psychology chapter. ACT* is thus a hybrid model: it specifies how multiple memory systems interact and how explicit knowledge is represented. A proposition is the smallest unit of knowledge that can be verified, that is, proven either true or false. Propositional networks allow for a greater variety of relationships among concepts, including actions. Figure 7.11 shows a propositional network that might underlie someone's understanding of his or her friend Bob. Each ellipse denotes a proposition. The proposition is defined by the arrow links that radiate away from it and point to concepts. There are different links that stand for different parts of the proposition. An **agent link** specifies the subject of the proposition, the one performing some action. An **object link** denotes the

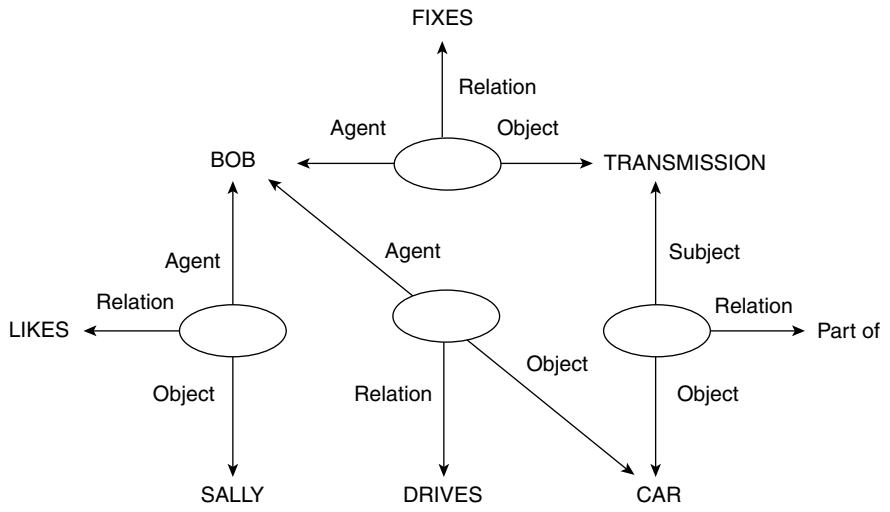


Figure 7.11 An example of a propositional network (after Anderson, 1995)

object or thing to which the action is directed. The **relation link** characterizes the relation between agent and object. The proposition "Bob likes Sally" would thus have "Bob" as its agent, "Sally" as the object, and "likes" as the relation.

Anderson's ACT* model can also account for the specific memories each of us has as part of our experience, episodic knowledge, and traditional factual world knowledge. His model does this via its creation of two classes of nodes. A **type** node corresponds to an entire category. The node that corresponds to "dogs" in general is a type node. **Token** nodes correspond to specific instances or specific items within a category. The node that represents "Fido," where Fido is the name of your dog, is a token node. In this way, the network can code for general world knowledge as well as personal memories.

Evaluating Semantic Networks

Semantic network models can be implemented as computer simulations and then put to the test to see if they account for human-like like memory processes. This has been done, and although many of the simulations can reproduce aspects of performance in limited domains, there is no associative network that successfully simulates all of human memory.

A phenomenon that presents a challenge to these network models is called the **T.O.T. phenomenon**, the acronym standing for "tip of the tongue." Try to

recall the name of the navigational device used by sailors that measures the positions of the stars. You may find yourself frustrated in trying to recall the word, as many people can come up with the first letter and the number of syllables but not the word itself (Brown, 1991). A person who gets stuck trying to remember the word is, in a manner of speaking, in the vicinity of the node or network that codes for the word, because he or she has some knowledge of the word. He or she has also presumably activated nodes that are in proximity to that of the concept in question, because he or she is devoting a great deal of time and effort to recalling the word. So why doesn't spreading activation activate the appropriate node? Semantic networks cannot easily explain these sorts of retrieval blocks. By the way, if you couldn't get it, the word is "sextant."

The phenomenon opposite to the T.O.T. phenomenon is the situation in which we can successfully retrieve an item from memory despite the fact that there are no close connections between retrieval cues and the target. Reisberg (2001) gives a good example when he asks us to think of the node that represents "water." Because water is an eminently familiar concept, it has connections to many others. The water node thus ought to have multiple links that radiate outward toward other nodes—a high **degree of fan**. In contrast, the concept "xylophone" has a low degree of fan, at least for most people, as we don't often think about this musical instrument. Assuming that the water node has links to 100 other nodes, probably a conservative estimate, then spreading activation would, in just one step, activate all 100. Assuming that each of these 100 nodes is, in turn, connected to 100 other nodes, then 10,000 (100×100) nodes would be activated in a memory search after only two steps. The number of activated nodes increases tremendously with increasing numbers of steps. This process is inefficient from a computational standpoint. It is also counterintuitive. It certainly doesn't feel as if we think of a million or more things when we are trying to remember the capital of Texas.

A solution to the problem of excessive activation would be the implementation of an inhibitory network. Here, the links are inhibitory, so activation of a given node causes it to dampen the activity of its neighbors. In this model the greater a node's activity, the more it will inhibit the nodes to which it is immediately linked. Nodes that receive several inputs are more highly activated and thus stand out in the crowd, since they can suppress their neighbors more powerfully. Concepts with multiple inputs are probably the ones we are searching for, because these input pathways are what are activated during the retrieval process. To go back to our earlier example, if we are trying to remember "France," this node ought to stand out among the others in an inhibitory network. This is because it is activated in two ways: on the basis of the "commonly taught languages" path, and on the basis of the "fought with Germany"

pathway. Inhibitory networks thus dampen activity in nodes that are unlikely candidates in a memory search, and increase activity in nodes that are likely candidates.

One of the nice things about semantic networks is that the search for an item takes place automatically. All one needs to do is hear a question and the answer seems to pop into one's head. The relative ease with which we recall, at least much of the time, is in accord with the idea of spreading activation. The question triggers nodes that are near the desired node; activation then spreads across the network until the target item is sufficiently activated. But, as we all know, recall is not always this easy. Sometimes, as in the T.O.T. phenomenon, we must expend effort in trying to remember something. If somebody asked you what you did on your last birthday and it didn't immediately come to mind, you might engage in a **guided search**—one governed by intelligence and reasoning. You would think of your friends and members of your family who might have been at the party, what kinds of presents you received, and where the party was held. All of these could serve as retrieval cues or associated nodes, and as such are elements in the network. But they weren't triggered automatically. Instead, you had to willfully and deliberately reconstruct what happened on your last birthday, based on what you know of your past birthdays in general. This type of **reconstructive memory** is used quite often and has been shown to be influenced by environmental cues, such as how the question is phrased (Loftus, 1979). But the important point here is that reconstructive memory constitutes a separate process of retrieving items, one that does not rely on spreading activation and the inherent, automatic characteristics of the network.

Overall Evaluation of the Network Approach

The network approach has received its share of criticism. Criticisms leveled against connectionism are that its architectures are too simple and that its learning rules are inadequate. Semantic and propositional networks have their own problems related to organization and the proper role of inhibition. Having said all this, the network approach is still a strong perspective, both in terms of its theoretical assumptions and methodology. Connectionist networks are good at simulating learning and have biological plausibility, at least in the gross sense of the term. Semantic and propositional networks probably come the closest to representing certain features of the way knowledge is represented in memory.

The network approach poses a number of challenges to the classic information processing view. Several differences need to be reconciled. These have to

do with the numbers and sizes of processing units, and include the argument of serial processing vs. parallel processing, as well as the argument of knowledge based vs. behavior based approaches. What is needed then is a way of integrating the opposing views. New computers with parallel processing architectures are being developed. These machines may yield new insights that will help to bridge the gap between the classic and connectionist perspectives. Another way to bridge this gap would be to have network researchers working with practitioners of other cognitive science disciplines in model building. Network models seem to constitute a "middle ground" between the actual biological structures studied in neuroscience and the more abstract representations of cognition and artificial intelligence. Model building that represents a cooperation between these disciplines might resolve some of the discrepancies.

In Depth: NETtalk

NETtalk is an ANN designed to read written English (Sejnowski & Rosenberg, 1986). It is presented with written letters of the alphabet. Its output is the correct pronunciation of the sounds represented by the letters, which is then fed to a speech synthesizer for production of the sounds. Unlike programs that existed at the time it was developed, NETtalk actually learned to make correct pronunciations after it had been supplied a given number of examples. Although the network could not understand what it was reading, it serves as a good demonstration of what networks of its kind are capable of and how they can serve as models of human learning.

The NETtalk system consists of three layers. The input layer has seven groups, each containing 29 individual units. The overall activity of these 29 units specifies one letter. NETtalk thus processes seven letters at a time. It focuses, though, on the fourth and middle letters of these seven. It is these target letters that the network attempts to pronounce. The other surrounding letters serve as context and to help disambiguate the correct pronunciation, due to the fact that in English, the sound of any given letter depends heavily on the adjacent letters. Figure 7.12 shows a diagrammatic representation of the network.

The input units next connect to a hidden layer of 80 nodes, which effect a partial recoding of the input data. These in turn connect to 26 nodes in the output layer, whose pattern of activation represents the system's initial response to the letters. This response is in effect the network's first guess at pronouncing the target letter. The guessed pronunciation is then compared to the correct response, specified by a teacher. Weights are then adjusted using the back-propagation algorithm. Finally, the "window" or read-head that viewed

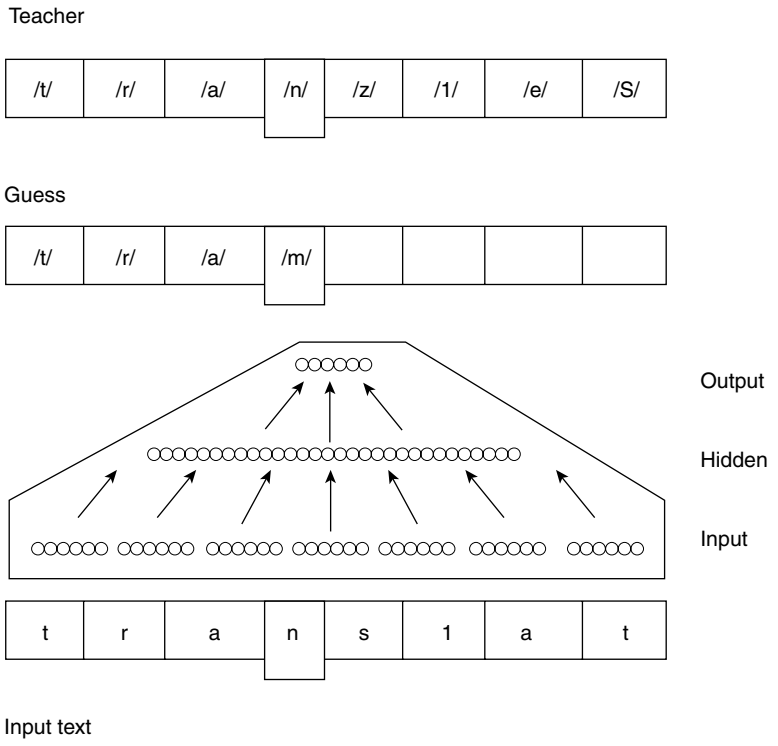


Figure 7.12 A diagrammatic representation of the NETtalk network, capable of converting written English text into speech. Sejnowski, T., & Rosenberg, C. (1986)

Source: *NETtalk: A parallel network that learns to read aloud.* (Technical Report JHU/EEC–86/01.) Baltimore: John Hopkins University Press. Reprinted with permission.

the seven letters at a time would move one letter to the right, to process the next letter in the text. The total number of weighted links that directly connect one node to another in NETtalk is 18,829, so this is a fairly large network by most standards.

How does NETtalk perform? After fifty passes through a sample text, which corresponded to about 250,000 letter-to-pronunciation pairings, it achieved a 95 percent accuracy rate. When the network was then tested on a new sample text without further training, it performed at 78 percent accuracy. This relatively impressive performance on the new text demonstrates the network's capacity for generalization; the network had extended its abilities to new words and letter combinations. NETtalk was also found to perform well after simulated damage. The simulated damage was in the form of changes made to

the weights: every weight was changed to a new random value that was within some prespecified range of the old value. When the randomized values were less than 65 percent of the average, performance was hardly affected—showing that the network is resistant to degradation.

Minds On Exercise: Free Association

In this exercise, you will be presented with some words and asked to write down whatever the words make you think of. Don't censor or inhibit your thoughts; just write down whatever comes to mind. This technique is called **free association** and is used by psychoanalytic therapists to assess the contents of the unconscious mind. We will be using free association to explore the structure of your long-term memory.

To begin, take out a piece of paper and a pen. Number three lists from one to ten. At the top of the first list write the word "ocean." Now take a minute to write down the first ten words that come to mind. When you are finished, write down the word "fish" at the top of the second list and again take a minute to jot down the first ten words that pass through your mind. Finally, write down "toothbrush" at the top of the third list, and do the same.

You may have noticed that the words "ocean" and "fish" are semantically related, and that "toothbrush" is somewhat unrelated to either. Did your first two lists contain any of the same words? Did you find it easier to think of words that had to do with "fish"? If so, that was probably due to priming. Was it more difficult to jot down words you associated with "toothbrush"? Why?

Turn the page over and sketch out a semantic network for each of the words. Represent each word as a node and draw connections between it and related nodes. Group related words by placing them close together. What is the structure of these networks? Is there any hierarchical organization? Were there any relationships between nodes that don't conform to the kinds of relationships that were discussed in this chapter, such as subject, object, or relation? Did certain nodes cluster together? How were the nodes in a cluster related? Compare one of your networks to that of a classmate. In what ways are they the same? In what ways are they different? Do you think different networks reflect different thinking styles or personality traits?

Food for Thought: Discussion Questions

1. Sketch out an ANN that performs arithmetic, reasons logically, or carries out some other mental process not discussed in this chapter. What is the input to this system? What functions do the input, hidden, and output layers perform? What is the output? How would the network learn?

2. Does a neural network have to be limited to three layers? Design one that has additional layers. What would the roles of these layers be?
3. We discussed several ways, such as the back-propagation algorithm, in which a neural network can learn. Can you come up with other learning algorithms?
4. Research on semantic networks shows that concepts cluster together based on shared characteristics. These include possession or belongingness, as found in the hierarchical model. In what other ways might concepts cluster together in a semantic network model? One approach to this problem is to free associate and then see what the relations are between successive words.
5. There are several different types of relation links that connect nodes in propositional networks. A relation can be an action, such as "chased." Can you think of others? Is there a problem with having so many possible relations?
6. Draw the propositional network that would represent a complex sentence such as "My friends hiked to the bottom of the Grand Canyon in August." How does one code for linguistic elements such as "of" or "the"? For pronouns? Adverbs?

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Anderson, J. (2000). *Learning and memory: An integrated approach* (2nd ed.). New York: John Wiley.

Hebb, D. O. (1949). *The organization of behavior*. New York: John Wiley.

Gurney, K. (1997). *An introduction to neural networks*. London: UCL Press.

Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA: MIT Press.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (Eds.) (1986). *Parallel distributed processing. Volume I: Foundations*. Cambridge, MA: MIT Press.

8

The Evolutionary Approach: Change Over Time

"I have called this principle, by which each slight variation, if useful, is preserved, by the term of natural selection."

—Charles Darwin, 1859

The Evolutionary View

In this chapter, we will examine the evolutionary approach and its relation to cognitive science from three perspectives. The first perspective, **evolutionary psychology** (EP), is concerned with how the human mind came into existence. It attempts to describe the forces in our ancestral past that could have given rise to mental structures. The second perspective, evolutionary computing (and the related discipline of artificial life), emerged from areas of study in artificial intelligence. They look at evolutionary processes as a form of adaptive problem solvings and can help us to understand human mental functioning. The third perspective, called Neural Darwinism, has its roots in neuroscience and sees evolutionary forces as being behind the creation of neural circuits. These circuits underlie our capacity to learn from and respond to the world.

Other approaches in cognitive science have adopted and used evolution. In the psychology chapter we saw functionalists asking the same questions that evolutionary psychologists ask. These functionalists wondered why we have certain mental abilities, to what purpose is a given ability, and how a specific mental ability may have come to be. Evolutionary principles are used also in contemporary robotics, in which generations of robots are bred to produce robot offspring that are best adapted to the tasks for which they've been designed.

A number of important themes are touched on in this chapter. Evolutionary psychologists argue that many of our cognitive mechanisms are innate. This argument echoes the Nativist position that we are born with rather than acquire knowledge. Evolutionary psychologists also argue that the mind consists of a set of special-purpose devices or evolved psychological mechanisms that are activated only within specific environmental contexts. This runs counter to a tenet of artificial intelligence that states that the mind is a general-purpose computing device, capable of solving virtually all kinds of problems equally well. The fields of evolutionary computing and Neural Darwinism postulate that evolutionary forces can explain the operation and formation of mental processes within the lifespan of a given individual. This is in contrast to the classical notion of evolutionary forces operating at the species level over much longer stretches of time.

Evolutionary Psychology

We humans don't look like much. We are thin-skinned—and almost gawky-looking when we stand upright. To a passing predator, we must seem like a particularly easy kill. We have no armor, fangs, or claws with which to defend ourselves. Nor are we particularly strong or fast in comparison to other animals. So what has enabled humans to survive and thrive in a harsh world? The answer, in a word, is our minds. *Homo sapiens* possesses the most sophisticated mental faculties of any species on earth. Our capacities to think, to use language, and to solve problems outstrip those of any other animal. It is our minds that set us apart. An advanced mind allows for planning, communication, social cooperation, and the development of technology, which give us a tremendous competitive survival advantage.

But how did the human mind come into existence? What processes formed it? These are some of the questions asked by the evolutionary approach. In evolutionary theory, changes in an environment give rise to corresponding physical changes in an animal species that better enable the members of that

species to survive. Giraffes, to use an often-cited example of this kind of adaptation to an environment, have over many generations developed longer necks to better enable them to get at treetop vegetation. Similarly, we can consider the mind to be such a body structure—comparable to the giraffe's neck—and one that has been shaped by evolutionary processes. EP not only attempts to describe these forces that are believed to have fashioned the human mind, but is also concerned with understanding the mind's organization and function, and with how interaction with an environment produces observable behaviors.

Evolutionary psychologists face a fundamental problem. They need to understand the selection forces that gave rise to mental structures, but cannot know precisely what these were because they correspond to events that were taking place far back in human prehistory. It is believed that the period of time during which many human psychological mechanisms evolved was the Pleistocene era, which began approximately two million years ago. This period is referred to as the **Environment of Evolutionary Adaptation** (EEA) (Bowlby, 1967). Because we cannot go back in time to study what occurred during this period, we must make inferences about the physical and social environment of that time, based on evidence of various types. Buss (1999) describes the types of evidence used to test evolutionary hypotheses. These include archeological records, data from studying hunter-gatherer societies, systematic observation of human behavior, and self-reports in the form of interviews and questionnaires. EP also relies on a variety of methods for the testing of its hypotheses. Among these are comparisons across species, comparisons of males and females within a species, and experimentation (Buss, 1999). The evolutionary approach thus adopts a wide variety of methods, rather than rely exclusively on the scientific method.

A Little Background: Natural Selection and Genetics

Before delving specifically into the evolution of mental structures, it is helpful to review some basic concepts of biological evolution. The biologist Charles Darwin (Figure 8.1), in his classic work *On the Origin of Species* (1859), outlines the basic principles of his **theory of natural selection**, which is a description of the process by which animal species change over time. Darwin noticed that animals vary (within a species as well as from species to species) in all sorts of ways—they have longer or shorter legs, bigger or smaller ears, or different coloration. The fact that animals differ in their physical traits is known as **variation**. Second, parent organisms pass on some of their gene-based

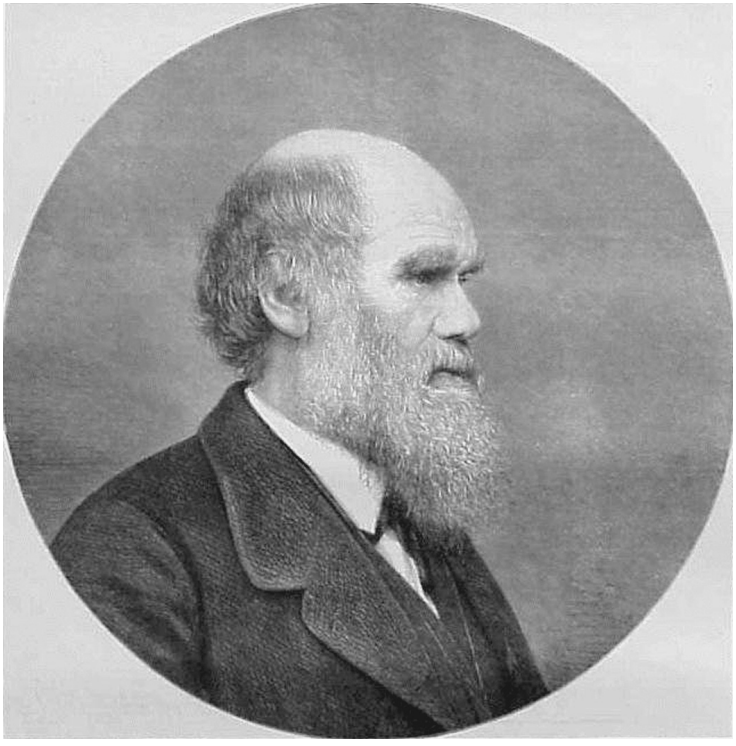


Figure 8.1 The biologist Charles Darwin (1809–1882) proposed the theory of natural selection

characteristics to their offspring. This process is called **inheritance**. A change in environmental conditions can make it more likely that animals that possess a specific trait will survive. Those that do not possess this trait will perhaps die. This process, whereby a particular attribute or attributes promotes survival under altered conditions, is known as **selection**.

Variation, inheritance, and selection together are the vital ingredients of species change. Looking at the fossil record, we can see that there had been alterations in the foot structure of horse-like animals. Early progenitors of the horse had feet with several toes. Figure 8.2 shows that, over time, these toes became fused and the hooves of modern-day horses came into being. Originally, the members of this progenitor species varied with respect to toe structure. Some animals had toes that were closer to each other and less

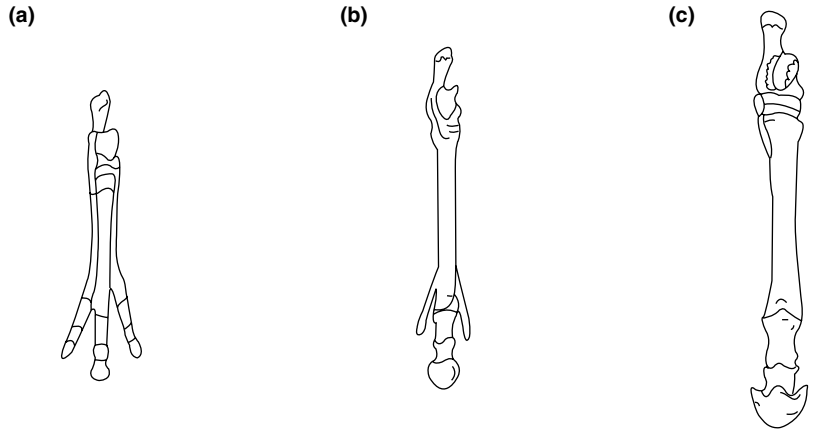


Figure 8.2 Evidence from the fossil record showing the evolution of the hoof from a toed foot in horse species

distinct anatomically. These animals were perhaps better able to run in an emerging grassland environment, as their feet were likely to have given them better support and locomotive power on flat terrain. Presumably these horses ran faster and so were able to escape predators more easily. They were more likely to survive and pass on their adaptive trait to future generations.

Natural selection is concerned with survival in the strictest sense: whether an organism lives or dies. Animals that possess an adaptive trait are better equipped to deal with a changed or changing environment. They are better at escaping predators, locating food, keeping warm or keeping cool, and so are likely to survive. **Sexual selection**, in contrast, refers to reproductive success—the ability to attract a mate and thereby produce offspring. Some traits that are selected for naturally are not always also selected for sexually. Females of some beetle species consider the presence of large horns on males of the species to be attractive. The males with these large horns mate more often and pass this trait to future generations more often. However, very large horns are also an impediment to survival. If they are too big, the males cannot easily

escape predators such as birds. In this case, the optimum horn size represents a compromise between the two types of selection trading off against each other.

Evolved Psychological Mechanisms

Evolutionary biology focuses on all aspects of an animal's makeup. It is concerned with how selection forces came to shape the individual parts of an organism and how these parts became integrated. The scope of EP is narrower—it examines the impact that these forces have had on specifically psychological mechanisms only. Whereas evolutionary biology looks at organs and organ systems, EP looks only at a single organ—the brain. This perspective views the mind as a collection of psychological mechanisms that have evolved and attempts to describe how each came to be and how each operates.

The traditional view in cognitive psychology is that the mind is a **general-purpose processor**. This processor can be "fed" a problem in any format and come up with a solution by applying a known set of rules. Modern computers operate in this way. The advantage of this process is that it can solve all problems of a given kind assuming the correct values of the problem's variables are mapped onto the algorithm designed to solve the problem correctly. The evolutionary perspective adopts a very different view. It argues that the mind is not general in any way, but instead a grab bag of many specific capacities. Cosmides and Tooby (1992) refer to the mind as a "Swiss army knife." In this analogy, each part of the knife unfolds to solve a different problem: the corkscrew takes out corks, the scissors cut paper, and so on. Likewise, each mental mechanism is assumed to have evolved in response to the need to solve a specific problem of adaptation and is activated by the contextual information of that problem alone. Evolved psychological mechanisms are examples of modules. The general characteristics of modules are discussed in Chapter 4 (The Cognitive Approach I).

Buss (1999) defines an **evolved psychological mechanism** as one of a set of processes inside an organism, each characterized by the following six properties:

1. An evolved psychological mechanism exists because it contributed to the solution of a specific problem of survival or reproduction (the mechanism is the product of natural or sexual selection). Arachnophobia, the fear of spiders, is an example of an evolved psychological mechanism (see Figure 8.3). Arachnophobia, as a mechanism, solves a very specific

survival problem: how to avoid being bitten by a poisonous spider. Humans who harbored this fear avoided spiders, were not bitten, and so were more likely to survive and pass the trait on to future generations.

2. Evolved psychological mechanisms are designed to take in only a narrow band of information. In biological evolution an organ such as the eye can only process a limited range of inputs—in the case of the eye, only those corresponding to the visible light spectrum. An eye is not capable of processing information that is, in effect, outside this range, for example, information that corresponds to the infrared or ultraviolet part of the electromagnetic spectrum. Similarly, psychological mechanisms are triggered by and can only process very particular types of input. As part of the psychological mechanism just described, only the image of a spider with its body parts and eight legs will elicit this particular fear—not an image of, for example, another creature, such as an ant.

3. The input of an evolved psychological mechanism communicates to an organism the particular adaptive problem it is facing. When confronted with a spider, an individual knows that he or she is dealing with a survival problem: the bite of the spider could be fatal. The person in question understands the spider does not pose some other adaptive problem, say, one related to food selection or the choice of a mating partner.

4. The input of an evolved psychological mechanism is transformed by decision rules into output. The image of a spider automatically triggers a response, such as avoiding the spider. This and similar responses can be thought of as decision rules that are in the form of "if-then" statements: if I see a spider, then I should run away.

5. The output of an evolved psychological mechanism can be physiological activity, information that will serve as input for other psychological mechanisms, or a behavior. Arachnophobes, upon seeing a spider, may become physiologically aroused. Their heart rate and respiration rate may increase, and they may begin to perspire. They may also use the information about the spider to help them to decide on their next course of action. If the spider is on the far side of a room, it makes sense to run away; but if the spider is on their shoulder, it might make more sense to brush it off. The decision-making process is another mechanism that operates by drawing on the information that has been received from the

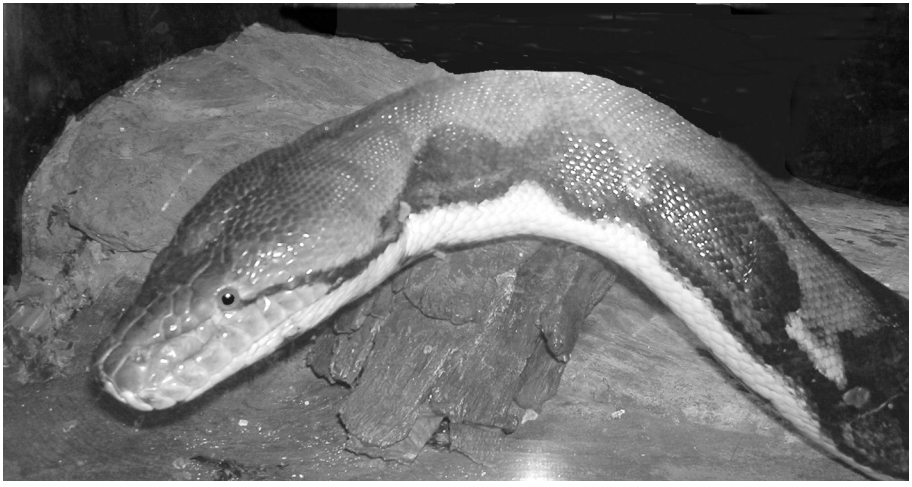


Figure 8.3 A reticulated python. Fear of snakes such as this one and of other animals, such as spiders, may be the result of an evolved psychological mechanism.

arachnophobic mechanism. The ultimate outcome is the behavior itself: running away, freezing, or another response.

6. The output of an evolved psychological mechanism is directed toward the solution to a specific adaptive problem. The behavior that results from seeing a spider is intended to solve the problem of how to avoid being bitten. Running away removes the person from the problem, ensuring that the spider does not pose a threat.

To sum up, EP views the mind not as a "one size fits all" machine, but as a collection of functionally distinct mechanisms. Each of these mechanisms was created by evolutionary forces operating over vast periods of time. Each was expressly "designed" to solve a problem our ancestors faced during the EEA. It is important to keep this multiple-mechanism view in mind as we describe specific examples of cognitive processes. For every possible cognitive process, it is not the brain as a whole that is acting (according to this perspective), but a particular component of the brain. Thus, when we talk about the evolution of memory or of logic, we are talking about specific mechanisms that may constitute only a part of the brain's memory or logic systems.

Evolution and Cognitive Processes

In the following sections, we describe specific evolutionary hypotheses that pertain to several domains in cognition and the evidence that supports them (Gaulin & McBurney, 2001; Palmer & Palmer, 2002). These hypotheses concern cognitive processes only, which include categorization, memory, logical reasoning, judgment under uncertainty, and language. We also offer a discussion of cognitive differences between the sexes. The motivation that underlies this research is the desire to arrive at an evolutionary account of how these cognitive processes arose. In other words: Why is it that we categorize and remember things the way we do? Why are we so poor at reasoning logically and making judgments under uncertain conditions? How is it that human beings alone of all the species on earth have developed such a complex linguistic ability? How can we account for the subtle differences between the sexes with respect to spatial and verbal abilities?

Categorization

Research conducted by Eleanor Rosch shows that people form mental categories not in an "either-or" fashion, but in a continuously graded way (Rosch, 1973, 1975). Why should this be the case? Why do we not lump things together into one group on the basis of a single characteristic, for example, "red objects" or "curved objects"? The organization of concepts, it turns out, is governed by "fuzzy" categories, whereby a particular item can be more or less representative of the category. In this section we will examine why humans form concepts in this way. Recall the finding, from the chapter on networks, that a bird such as a robin is thought of as being more representative or prototypical of the category of birds than another type of bird, for example, the penguin. This is because robins tend to come closer to our idea of what a bird is—something that is small, sings, flies, and so on. Robins are thus more typical of "birds" than penguins, and participants in studies respond more quickly when they are asked to make judgments that have to do with robins—the so-called **typicality effect** (Rips, Shoben & Smith, 1973). See Figure 8.4.

Categories of the sort found in nature are almost always continuous: they contain items that span the entire spectrum of typicality: items that range from being very representative and prototypical to ones that are quite unrepresentative and perhaps easily mistaken for items that belong to another category. The categories that result from human categorization processes therefore mirror the distribution of natural objects in the world. This suggests an evolutionary process whereby we came to acquire concepts and a sense of their relatedness

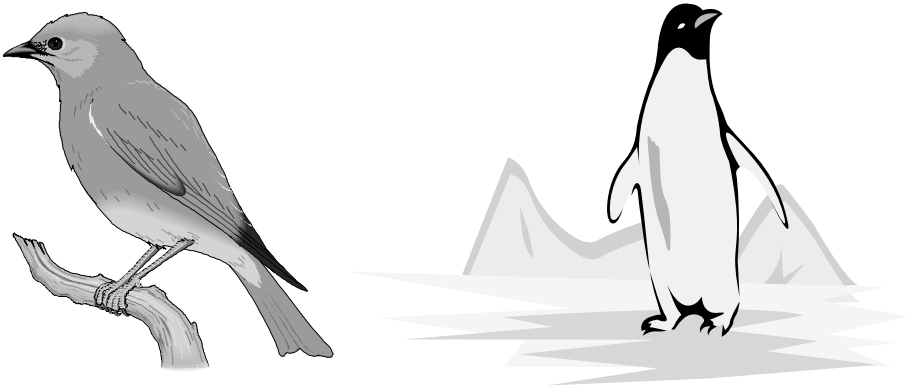


Figure 8.4 Which of these birds looks more like a bird? What does this tell you about your concept of bird?

on the basis of how the items that these concepts represented were ordered in the natural environment. Further evidence in support of this notion comes from Boster and d'Andrade (1989). They found that professional ornithologists as well as those who had little experience and/or knowledge of birds all grouped stuffed bird specimens in similar ways. Their groupings demonstrated typicality and hierarchical effects reflective of what exists in nature.

What purpose do typicality-based categories serve? They must have provided us with some sort of selective advantage, otherwise they would not have arisen. The advantage is this: if we know something about an item of which we have had experience, then it becomes possible to form judgments that have to do with related items (Rips, 1975). Imagine encountering for the first time a plant with four leaves and purple spots. After eating this plant you become ill. Several months later you stumble across a plant with three leaves and reddish-purple spots. Would you eat it? Probably not. There is a good chance that the two species of plant, because they are similar, share properties. Typicality categories thus allow us to capitalize on past experience. We generalize from what we know to what we don't know.

Memory

Other approaches in cognitive science tend to study basic memory processes. Cognitive psychologists, you will recall, are concerned with, among

other things, how information is transferred from short-term to long-term memory. Neuroscience can identify the brain structures that underlie this transfer. Connectionists can then create artificial networks to model the process. EP, on the other hand, cares less about memory function per se and more about the purposes our memories serve. EP researchers want to know why our memories are the way they are. Specifically, they want to know how the demands of our ancestral environment have shaped memory and have made our memories operate the way they do. In this section, we examine one aspect of memory: the relationship between the frequency of contact with an item in one's environment and recall.

Every day, we are bombarded with a tremendous amount of information. Only a small fraction of it is actually remembered. This is a good thing, as most of the information is trivial. For instance, can you remember the color of the shirt your best friend was wearing last Monday? It would make sense, then, for us to remember only the information that may be of some benefit to us, and to forget that which doesn't benefit us. But what constitutes beneficial information? What things should we be better at remembering? An evolutionary account avers that information that we encounter more often should be better remembered. The more often we encounter an item, the greater its relevance to our daily life and our survival. The less often we encounter an item, the smaller is its relevance to our survival. Remembering your roommate's name is important—because he can help you in a variety of ways. You would also be likely to hear your roommate's name quite often, as you see him practically every day. Remembering the name of your niece who lives in another country is less important to you, as you are far removed from interacting with her on a daily basis. Correspondingly, you would very rarely hear her name, assuming it is not the same name as that of someone else you know better.

One way to test this hypothesis is to compare the frequency of our encounters with specific items with our ability to remember them. Clearly, this relationship will be positive: the more frequently we encounter items, the better our memory of them. In addition, the relationship is described by a power function. For our purposes, we only need to know that a power function is a specific kind of mathematical relationship. This specific relationship is represented when we plot on a graph our ability to remember items versus how often we encounter them.

Anderson and Schooler (1991) tested this relationship. They examined several real-world sources of information and determined for each type of information the function for the frequency of occurrence and recall. That is, they determined the mathematical relationship specifying how many times a piece of information occurred and how accurately it was remembered. They used

newspaper headlines, utterances made by parents to their children, and names of authors of e-mails that were received by one of the researchers over an approximate four-year period. For all of these sources, the probability that a particular utterance or word would crop up a second time was a decreasing function of the number of days since its last appearance. For example, if the word "north" appeared in a newspaper headline yesterday, there would be a relatively good chance that it would appear again today, presumably because it was part of an ongoing news story. If "north" had been in the paper last week, the chances it would appear again today would be much lower. The researchers then measured memory for these items and found that the function that described probability of recall over time was in accordance with probability of occurrence. This shows that our memory for items is directly proportional to how often we encounter them in the environment. Our memories are better for items that we encounter more often and poorer for those we encounter less often. This finding supports the notion that our memories have evolved such that they retain items only to the extent that we need them. From an evolutionary point of view this makes sense. We should remember those things that we encounter frequently as it is very likely that they are of importance to us, and forget those things that we encounter infrequently, as it is likely that they are of not much use to our survival.

Logical Reasoning

Recall from the chapter on philosophy that deductive reasoning involves the application of logical rules to a set of premises. A conclusion is then derived. If the premises are true, then an application of the rules will always yield a correct conclusion. Deductive thought is thus a valid way of garnering new information about the world. Logic is used in other approaches to cognitive science. Connectionists use the rules of logic to describe how neural units interact. Researchers in artificial intelligence create computer programs that use logic to generate new information from existing data and to solve problems. In the evolutionary approach, investigators want to know why we as humans are so bad at thinking logically and to relate this deficiency to the kinds of social conditions our early ancestors may have faced.

If you have ever taken a course in logic, you may have wondered why the subject is so difficult. The rules of logic are straightforward. All one has to do is memorize the forms of these rules and then plug into them the relevant aspects of the problem. In this fashion, one can then apply the forms to any problem and be assured of obtaining a correct solution. Unfortunately, it is not as easy as it sounds. Most people find it difficult to reason logically. Interestingly,

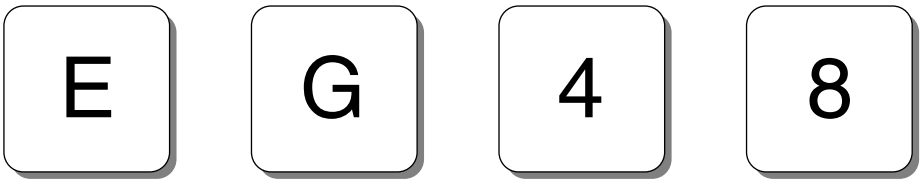


Figure 8.5 You have been hired as a clerk. Your job is to make sure that a set of documents is marked correctly, according to the following rule: "If the document has an E rating, then it must be marked code 4." You have been told that there are some errors in the way the documents have been coded, and that you need to find the errors. Each document has a letter rating on one side and a numerical code on the other. Here are four documents. Which document(s) do you need to turn over to check for errors?

Figure 8.6 You have been hired as a bouncer in a bar and you must enforce the following rule: "If a person is drinking vodka, then he must be over twenty years old." The cards depicted in the figure contain information about four people in the bar. One side of each card lists a person's age and the other side specifies what he or she is drinking. Which card(s) do you need to turn over to make sure that no one is breaking the law?

however, there is an exception. People can reason logically under certain conditions. The evolutionary approach can tell us why we are so bad at logical reasoning in general, and what the exception is.

The **Wason Selection Task** is designed to measure a person's logical thinking ability. Figure 8.5 gives an example of such a task. Take a minute now and try to solve this problem. You probably found that it was quite difficult. Next, try to solve a different version of the problem, shown in Figure 8.6. When you have finished—or if you are having trouble—consult Figure 8.7 for the correct solutions. If you are like most people, you found the "bouncer" version of the problem much easier. Why is this? Both versions have the same underlying logical structure and should be of equal difficulty.

According to Cosmides and Tooby (1992), the bouncer problem is easy because it involves **cheater-detection**, the ability to discern who has taken as

Solution to the "clerk" problem.
Rule: E-rated documents must be marked code 4.
Step 1: Check E-rated documents to see if they meet the coding criterion.
Step 2: Check any documents not matching the coding criterion (e.g., code 8) to see if they are E-rated.
No more checking necessary:
G ratings (paired with any code) could not violate the rule.
Code 4 documents do not have to be E-rated.

Solution to the "bouncer" problem.
Rule: Vodka drinkers must be over twenty years old.
Step 1: Check vodka drinkers to see if they meet the age criterion.
Step 2: Check anyone who does not meet the age criterion (e.g., 17 years old) to see if they are drinking vodka.
No more checking necessary:
Pepsi drinking (at any age) could not violate the rule.
People over twenty do not have to drink vodka.

Figure 8.7 Solutions to the "clerk" and "bouncer" problems

his own something he does not deserve. They argue that during the EEA, human societies were characterized by **reciprocal altruism**, the sharing of hard-won resources among group members. In this kind of social cooperative it is important to detect "freeloaders," those who take more than their fair share. Cosmides and Tooby believe that this social environment shaped our reasoning abilities such that when we are presented with a logic problem that involves cheating, we can solve it. If a logic puzzle possesses the element of a cheater, then the evolved psychological mechanism for cheater-detection is activated. This mechanism "understands" the rules of logic and can apply them, but only in this specific context. Notice that this understanding stands in stark contrast to that which maintains that the mind is a general-purpose processor that can solve logic problems of any sort.

There is abundant experimental evidence in support of the cheater-detection model of reasoning. One study comes from Gigerenzer and Hug (1992). They presented participants with two Wason Selection tasks, both of which entailed a social contract of the sort: "If you stay overnight in a mountain shelter, you must help out by carrying up some firewood to add to the supply." Both of these tasks were roughly equivalent with respect to content and logical structure. The only real difference between the two was that one of the tasks involved cheating, while

the other did not. The results showed that 78 to 90 percent of the participants were able to solve the cheating version. Only about half as many participants were able to solve the non-cheating version. This shows that the framing of a logic problem as a social contract alone is not enough to elicit problem solving skill; it is elicited by a social contract that specifically involves cheating. In another study, it was discovered that even children a little older than two years of age seem to have this cheater detecting capacity (Cummins, 1999).

Judgment Under Uncertainty

In many situations in life, we have to make a decision in the absence of complete information. What are the chances that Mary will go out with me? What is the likelihood that I will make the high school swim team? As is the case with logical reasoning, most people are quite poor at solving problems of this type—problems that involve probability (Nisbett & Ross, 1980; Tversky & Kahneman, 1974). Why is this? One reason is that humans facing these kinds of problems rely on heuristics. A **heuristic** is a mental "rule of thumb" or strategy that is a fast and easy way of solving a problem. The problem with heuristics is that they aren't always right. Heuristics can lead us to commit ourselves to **fallacies**, fundamental misunderstandings of statistical rules. The role that evolution has played in shaping our thinking in this situation is discussed below.

Tversky and Kahneman (1974) were the first to study heuristics and fallacies. In one study, they gave participants a description of a person. Participants were also told that this person belonged to a set of 100 people. There were two conditions in the experiment. In the first, participants were told that the set consisted of 30 lawyers and 70 engineers. In the second, they were told that the set contained 70 lawyers and 30 engineers. Here is a sample description: "Jack is a 45 year old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies, including home carpentry, sailing, and mathematical puzzles." Is Jack a lawyer or an engineer? The participants answered overwhelmingly that Jack was an engineer, even in the condition in which engineers made up only 30% of the set. Tversky and Kahneman argue that this kind of error is due to a **representativeness heuristic**, the tendency to judge an item on the basis of its perceived similarity to other items. Because the description of Jack fits the stereotype of the engineer, we consider it very likely that he is one. The representativeness heuristic in this case is accompanied by the **base-rate fallacy**, ignorance of the base rates that define the set of 100 people.

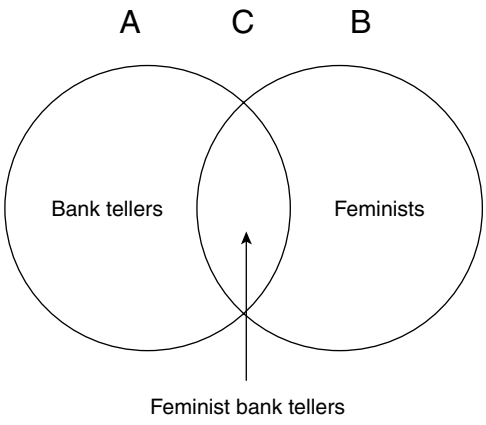


Figure 8.8 A Venn diagram illustrating the conjunction rule. Circle A represents all bank tellers. Circle B represents all feminists. The region C, the region of overlap between the two circles, represents feminist bank tellers and will always be smaller than the regions that correspond to the number of bank tellers or feminists, considered separately

Here is another example of our difficulty with making judgments under uncertainty (Tversky & Kahneman, 1974). Consider the following description: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. In college, she was involved in several social issues, including the environment, the peace campaign, and the anti-nuclear campaign." Which of these statements do you think is more likely: "Linda is a bank teller," or "Linda is a bank teller and is active in the feminist movement"? Most people overwhelmingly choose the latter sentence, even though it cannot possibly be the more likely of the two, since there are always going to be more tellers than feminist tellers (see Figure 8.8). Participants again rely on the representativeness of Linda's description, which fits the stereotype of the feminist. In this instance, participants ignore the conjunction rule, which states that the probability of encountering those who are both feminists and tellers is lower than the probability of encountering those who are one or the other. The making of this error is known as the **conjunction fallacy**.

So, is it fair to say that humans cannot think in terms of probability or uncertainty? From an evolutionary standpoint, one would guess that this is not a fair statement, as uncertainty is an inherent characteristic of the natural world. Gigerenzer and Hoffrage (1995) argue that the difficulties that human beings have in this area, which are documented in the literature, stem not from a fundamental inability on our part, but from the way that these types of

problems are usually structured. Many of these problems entail probabilities that are expressed as fractions or percentages, rather than as frequencies, expressed as numbers of occurrences. Probabilities are more complex numerically as they require dividing one number by another. They also are the rather recent invention of mathematicians. We encounter frequency data, on the other hand, as part of our daily existence. We can think easily of how many times we've eaten today or how many games the New York Yankees have won this season.

It makes sense, then, that the psychological mechanism that underlies uncertain reasoning should be sensitive to frequency data. It turns out this is indeed the case. Gigerenzer and Hoffrage (1995) assessed a number of uncertainty problems and converted them from a percentage to a frequency format. They found that roughly three times as many participants answered correctly. Their finding, similar to the findings of the logical reasoning study, illustrated earlier, is another example of how problem solving is narrowly tuned to a particular type of information. This attunement to a particular type of information is the second major characteristic of evolved psychological mechanisms and again runs contrary to the notion of the mind as a general-purpose processor—at least within these problem-solving domains.

Yet another example of poor probabilistic reasoning is **the gambler's fallacy**. Do you think it is more likely that someone who has played the slot machines for a longer period of time will win? Do you think that a fair coin that has come up "heads" eight times in a row is more likely to come up "tails" on the next toss? If you answered yes to either of these questions, then you have been taken in by the gambler's fallacy. It is the belief that the probability of an event is influenced by how often it has occurred in the past. If a particular outcome has turned up quite often in the recent past, people are apt to believe that there is an increased likelihood that another outcome will occur. In reality, the probability of each event in these cases is independent. The chances of winning at slots the one-hundredth time are exactly the same as for the first. The chance of getting a "heads" in a fair coin flip is 0.50 on every flip.

Why then do we persist in holding such beliefs? One explanation is that independent outcomes are typically the product of artificially engineered devices that are not encountered in the natural world, such as dice or roulette wheels (Pinker, 1997). In nature, the past often *is* the best predictor of the future. If it was sunny yesterday, there is a good chance that it will be sunny today. If a herd of mammoth migrated through a particular region last year at this time, chances are that they will do it again this year. Natural events tend to have regular time courses and shared causes, and therefore some knowledge of the way an event unfolded in the past is informative about the way that similar events will unfold in the future.

Language

Language is so important in cognitive science that it is a separate field of study. Linguists study the organizational rules of language and how language is acquired. They also construct models of language processing. From an evolutionary standpoint the focus again is not so much on processing but on how language came to be. Because language possesses a social function, evolutionary researchers want to know what aspects of early human social life might have given rise to its development.

One of the most obvious things that sets *Homo sapiens* apart from other species is language. Other animals can communicate and primates can be taught rudimentary linguistic skills, but nowhere else in the animal kingdom do we see such a sophisticated use of and widespread reliance on language as in the human species. The evolution of language is believed to have catalyzed the development of intelligence, social cooperation, and other phenomena, in the process of which human beings derived great benefit (see In Depth: A Stage Theory of Evolution). Researchers in this area have argued that language is an evolved trait and have provided several theoretical explanations for how it could have arisen. Non-evolutionary accounts of language are discussed in Chapter 9 (The Linguistic Approach).

Converging lines of evidence point to the fact that human language ability has evolved (Palmer & Palmer, 2002). First, languages share a number of universal commonalities (Hockett, 1960). Among these are meaningfulness, the ability of those who possess language to communicate about things not physically present, and the ability of those who possess it to produce an infinite number of new meanings via the combination of existing elements. All babies progress through the same developmental language stages, in the same order, and on roughly the same time schedule. Children also acquire the rules of language faster than they could possibly acquire them through formal learning. Finally, there are specific brain areas devoted to language. Most individuals are left hemisphere dominant with respect to language and have localized sites within this hemisphere that are devoted to linguistic comprehension and linguistic performance. These observations strongly imply that language capacity is prespecified and is present at birth. This in turn means that language capacity is genetically coded and has been shaped by selection forces that were operational at some time in our ancestral past.

Robin Dunbar (1996) has developed a fascinating theory of language evolution. He notes that there is in hominids a positive correlation between group size and the size of the neocortex. The neocortex underlies cognitive function and is quite large in humans. This brain region processes social information, for example, information that has to do with which members of a group are

our friends and which are our enemies. These kinds of social computations become more complex with increases in group size.

Most primates form alliances by grooming one another—by picking parasites and dirt from the fur of one another. This feels good to them and is a form of social bonding. But grooming, because it is a one-on-one activity, imposes a limit on how many individuals in a group can bond. Once a group becomes large enough, not every member of the group can be groomed. Language is also a form of social bonding and a way for individuals who possess this capacity to form alliances within groups. These alliances are forged primarily through the equivalent of gossip, wherein small groups communicate in both positive and negative ways about other group members. Because one can speak comfortably to a larger number of other individuals at once, language may have driven the formation of larger groups in early proto-human societies.

The coming into being of larger groups prompted further innovations, which were adaptive in nature. Larger groups are better able to defend themselves against predators in an open savanna environment. Language allows for complex, coordinated social behavior. Language would have facilitated hunting, foraging, childcare, and other crucial survival behaviors. The development of language may have received an additional boost from sexual selection. Assuming females recognize verbal skill is a rough indicator of intelligence, then they might have chosen to mate with males who were good "talkers," as their intelligence would have made them better partners.

Sex Differences in Cognition

Married couples may sometimes be aware of the existence of cognitive differences between the sexes. Perhaps one partner gets lost, while the other is forgetful. Research has shown that, in the human species, there are small but consistent cognitive differences between males and females. The evolutionary approach tries to provide explanation for them in terms of the difference in the selection pressures that were acting on the two sexes during the EEA.

Important cognitive sex differences are attributed to an early human **sexual division of labor**, according to which men hunted and women gathered (Silverman & Eals, 1992; Silverman & Phillips, 1998). The hunting process, which was apt to require navigation across large distances and novel terrain for days at a time, is believed to have selected for enhanced spatial abilities in men (Figure 8.9). Women, for whom communication with one another and with their children was a part of the gathering process, are believed to have developed superior verbal abilities (Figure 8.10). Indeed, such differences between

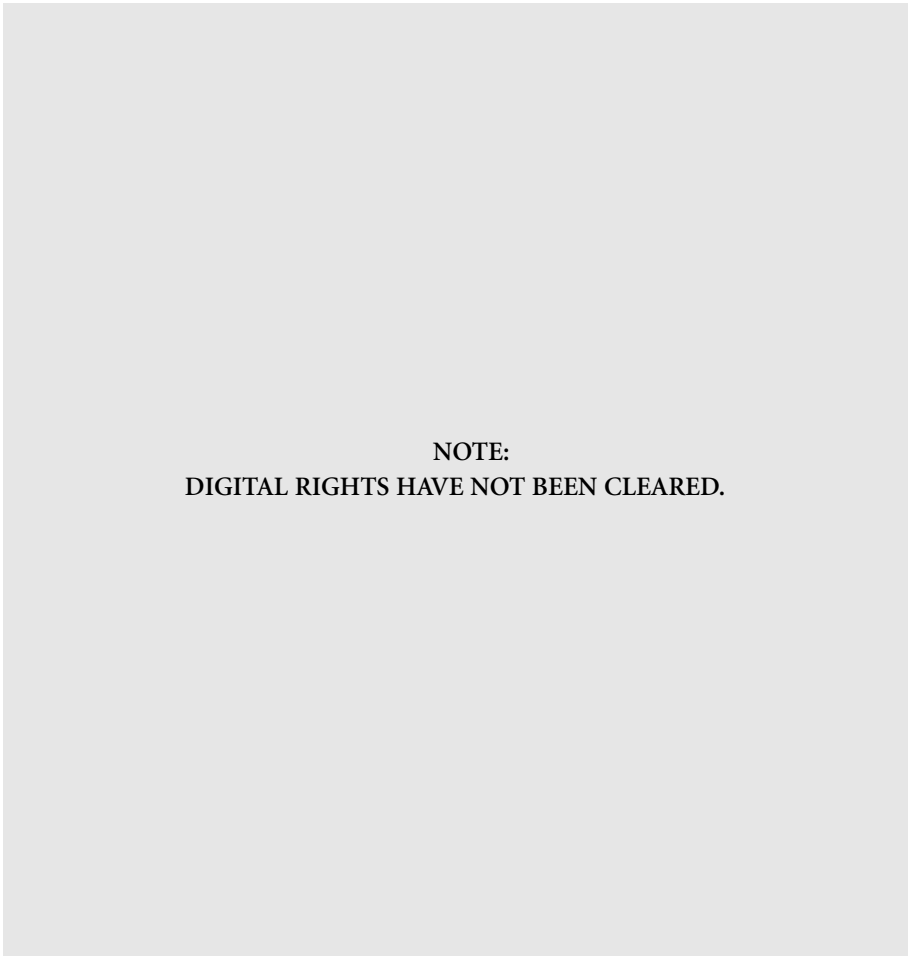


Figure 8.9 Men who are part of a traditional society hunting. What cognitive skills do you think hunting requires?

Source: Photo courtesy of AnthroPhoto.

the sexes are observable today. On average males outperform females on tests of spatial relations and image rotation (Levy & Heller, 1992). In these tests, a participant is usually asked to determine whether a letter of the alphabet, displayed in a variety of orientations, is the same as another. Females generally score higher than males on tests of verbal fluency, reading speed, reading comprehension, and spelling (Notman & Nadelson, 1991). One might also deduce that, as women were specialized at gathering, they should be adept at specific

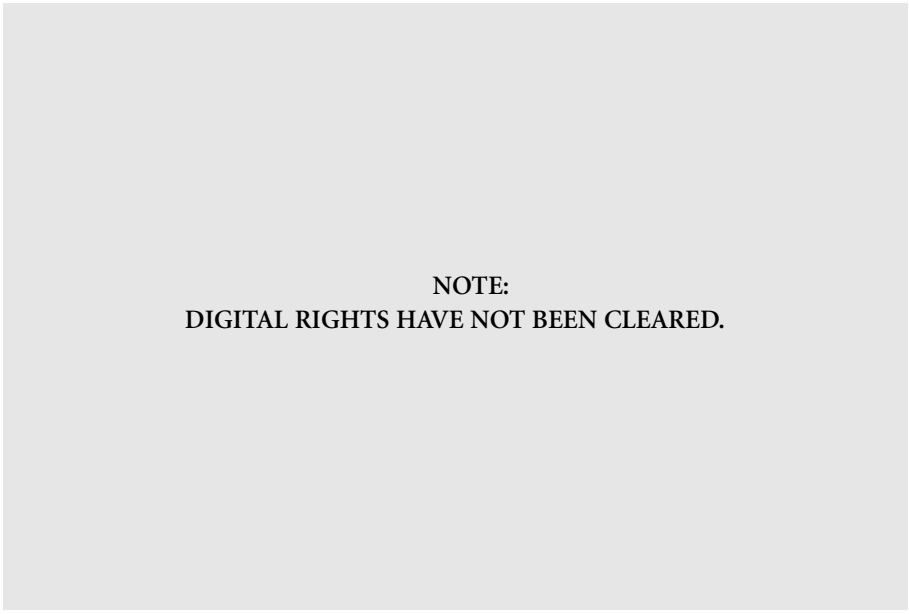


Figure 8.10 Women engaged in gathering. What cognitive skills make for a good gatherer?

Source: Photo courtesy of AnthroPhoto.

kinds of spatial tasks, namely, object location memory. Memory for object location is imperative to the task of gathering, as one would need to recall the locations of perhaps widely disseminated food sources. Silverman and Phillips (1998) discovered that women outperform men in their ability to remember the locations of objects, both natural and artifactual (see Figure 8.11). Their study suggests that our conceptions of spatial and verbal abilities are crude and that there are perhaps multiple domains within these abilities that may point the way to additional sex differences.

A second theory that accounts for the differences between men and women with respect to spatial and verbal abilities has been advanced. According to this hypothesis spatial ability became enhanced in males not as a result of hunting activity, but as a result of longstanding mate-seeking practices. The males of polygynous animal species must travel comparatively greater distances to meet up with and mate with females. Data indicate that the males of polygynous species—for example, the meadow vole—make fewer errors than females when given maze navigation tasks (Gaulin & FitzGerald, 1986, 1989). These sex

Figure 8.11 A stimulus set that can be used to test for object and location memory

differences disappear in the prairie vole, which is a monogamous species and neither sex of which ranges very far. It should be noted that there is nothing special about "maleness" or spatial ability. Rather, it is the environmental demands that are placed on one sex or the other of a particular species that are pivotal. In the brown-headed cowbird, there is a greater need for spatial problem solving among females, who show evidence of spatial ability that is superior to that of the males (Sherry et al., 1993). Among humans, it is possible that the enhanced spatial ability of males developed first as a result of mate-seeking and was later co-opted for hunting activities.

Evolutionary Computing

So far, we have been considering evolutionary processes in relation to how they may have shaped psychological mechanisms. However, an understanding of the process of evolution can also be applied to describing the way in which those mechanisms themselves function. In this section, we will detail how evolution is an analogy for some mental operations and can be used as a tool to help us to understand how the mind operates—not just how it came into being. That is, we will consider evolution as a process that is equivalent to some mental operations, rather than simply as a collection of historical events that may have given rise to these operations.

Recent years have seen the introduction of a new field of study that is known as **evolutionary computing** (EC). It is a collection of computational methods that have been modeled on the principles of biological evolution (Mitchell, 1996). EC serves three basic purposes. It solves real world problems, including problems having to do with financial forecasting, computer program design, and robot learning—the last of which is discussed in the chapter on robotics. EC is also used to model and further understanding of natural evolutionary systems that exist within fields such as economics, immunology, and ecology. Third, and most important for our purposes, it serves as a metaphor for the operation of human thought processes.

EC has three methodologies. These are genetic algorithms, evolutionary strategies, and evolutionary programming. Some genetic algorithms have been implemented specifically as artificial neural networks. Regardless of the methods used, all forms of EC share the same general approach to problem solving, which consists of several stages. These are:

1. The generation of possible solutions or "candidate solutions"—that is, a population of "individuals" who possess a range of features or characteristics.

2. An evaluation of the fitness of the solutions—of how well each solves the given problem.
3. A selection of the solutions that have the highest fitness values, based on some predetermined fitness threshold.
4. The production of a new generation of "offspring" from the solutions that have the highest fitness values. The offspring are made to "reproduce" via a crossover of "genetic material" from two or more "parents." These offspring can also undergo random mutations.
5. Steps 2–4 are repeated, until the optimum solution is obtained.

To illustrate this approach: one could use EC to design a new protein, perhaps to combat some disease. Proteins are strings of amino acids—there being 20 amino acids in all. If our plan was to engineer a protein that consisted of a string of 14 amino acids, then the total number of possible strings we could obtain—the total number of candidate solutions—would be 20^{14}. The number of candidate solution proteins would become reduced with each successive iteration of the program, until we ultimately ended up with a protein that possessed the desired characteristics.

EC cannot solve all problems well. For some types of problems, there are other computing techniques that are faster and more efficient. John Koza (1992) describes a genetic algorithm that was designed to solve a problem that involved the stacking of blocks. The program computed a solution—but one that required 2,319 block movements! The programmers had not specified that a good solution was one with few movements and that the best solution was the one with the fewest movements. This illustrates the importance of selection criteria in the identification of fitness values.

In general terms, the human approach to problem solving sometimes seems evolutionary. We are apt to generate a number of ideas, or solutions to a problem, initially. The common term for this is "brainstorming." We then evaluate those ideas on the basis of merit, perhaps using a variety of criteria. We select one idea or solution as being the best and then make additional selections in an attempt to refine the original selection. We can apply this method to the everyday scenario of deciding which kind of food to eat for lunch. One could imagine eating Chinese, Indian, or Italian. In the evaluation phase, we might decide against Chinese, because we had eaten it yesterday, and against Italian, because it is too expensive. Having decided on Indian food, we might then choose between two Indian restaurants, and select one, because it had better service. Evolution is then a metaphor that helps us to understand the human

decision-making process, wherein the generation of ideas is followed by the staged pruning of those ideas.

Artificial Life

Artificial life (A-Life) is the study of man-made systems that have been designed to behave in ways that simulate the behavior of natural living systems (Langton, 1989). A-Life researchers produce artificial "creatures" that inhabit virtual environments. They then study the behavior of these creatures and how they interact with each other and with aspects of their environments. What makes A-Life interesting and relevant to cognitive science is that, although the rules that govern the creatures are often quite simple, the emergent behavior of these creatures is complex and can be considered intelligent. We discuss A-Life in this chapter because these creatures learn and adapt via the use of evolutionary principles.

A good starting point for a discussion of A-life is Animat (Wilson, 1985). Animat is a single animal that lives in a simple environment. Its world consists of a square computer screen taken up by food items, which Animat can eat, and trees, which are obstacles Animat needs to navigate around. Animat can sense its environment; it is aware of the objects that are in its immediate vicinity. It can also act on its environment by moving one step at a time in one of the eight directions that are specified by the points of a compass (N, NE, E, SE, and so on).

Animat is initially endowed with a set of rules that govern its performance. These rules are in the form of classifiers, templates that match an environmental condition that Animat is subject to at a given moment to an action. Classifiers are really a modified version of if-then production rules. Each classifier has an associated strength or fitness. Classifiers with greater strength values are more likely to be utilized by Animat, as they promote Animat's survival. Animat's actions also have consequences. A good action is one that promotes its survival, such as its moving to a location where there is food. Good actions are rewarded by increases in the strengths of their classifiers. Neutral actions, ones that have no effect on the well-being of Animat, such as bumping into a tree, engender no change in classifier strength.

Animat is "born" with a collection of classifiers. These classifiers, along with their strength values, are randomly determined and as such give Animat no inherent advantage. These starting classifiers may be considered as equivalent to the genetic variation that exists among individuals or the members of a

species and that is shaped by selection forces. As Animat moves around its little world, classifiers are chosen and their strength values are altered on the basis of a system of reinforcement in such a way that survival-enhancing actions become more likely. However, this system alone is not enough to ensure Animat's survival. What Animat needs is a way of creating new classifiers that will better attune it to its world. This is where evolution comes in.

Wilson (1985) has equipped Animat with an evolutionary algorithm. Two classifiers that correspond to the same general type of action are selected. Individual components of these classifiers are then swapped. Segments of each classifier are randomly picked and exchanged, then used to create new classifier offspring. This is of course analogous to sexual reproduction in which genetic material from both parents is used to create children. Wilson, however, also allows for asexual reproduction. In this case, individual classifiers are cloned (reproduced in their entirety) to have random mutations.

The result of all this is that Animat's behavior becomes gradually more adaptive. Rather than modify the probabilities of its selections that represent its existing genetic endowment, it is able to generate entirely new actions, at least some of which will have beneficial results. It has been shown that these Animats, along with some more recent versions, learn quite rapidly. In one simulation, Animat was able to find food in only four steps, compared to a chance level performance in which it did the same in forty-one steps.

Other A-Life programs demonstrate greater complexity. Ackley and Littman (1992) have created an entire population of creatures. Some are herbivores that eat plants. Others are carnivores that hunt and eat herbivores. Over time the herbivores learned to "climb" trees to avoid predators, while the carnivores learned to cannibalize their fellow carnivores who died of starvation! Other emergent social behaviors of A-Life creatures include parasitism, symbiosis, and flocking.

A-Life simulations show that evolution-based software entities interacting in a diversified environment are capable of complex behavior. If we define intelligence as one's ability to modify one's behavior in an advantageous way over time in the pursuit of goals, then Animat and other A-Life creatures may be labeled as intelligent. There is, however, no generally accepted definition of intelligence; therefore the issue of whether such creatures possess it remains unresolved.

The field of A-Life is closely tied to the study of artificial intelligence. One of the key differences between the two fields is that A-Life attempts to model the behavior of the entire organism, rather than the functioning of the mind or some specific mental process (Beer, 1990). Because the behavior of A-Life creatures is reactive, they lack an "understanding" of their world; they have no means of representing knowledge. Another criticism of A-Life has to do with

the difference between simulation and realization. It is one thing to simulate life (some say), but another thing altogether to actually instantiate or create it in a true physical sense. Much of real life's complexity, exemplified by the ability to perceive and to move around, is given to A-Life creatures in the form of default abilities, in the absence of any real explanation as to how they may have arisen in these creatures.

Neural Darwinism

Gerald Edelman (1989) has developed a theory of **Neural Darwinism**; it applies the idea of evolutionary processes to neural learning. The theory is worth reviewing because it shows that the rudiments of natural selection, that is, selection and variation, are capable of forging neural circuits. Evolutionary forces may therefore be at work both at the level of thinking and problem solving, as suggested by EC studies, and at the level of neural circuit formation, which in fact underlies our thought processes.

In this view, new connections between cells are formed in three ways. First, there is the formation of connections and neural structures that occurs as part of embryonic development. These structures contain a great deal of variability and lay the groundwork for later learning. The result of this neural development is what is called a primary repertoire of neuronal groups. Second, the organism's eventual interaction with the environment prompts the differential modification of synaptic strengths, wherein strong connections among existing collections of neurons are forged. These modifications, which correspond to a selection of certain pathways over others, allow the organism to respond adaptively to its environment. The result is a secondary repertoire of neuronal groups. The third way in which connections are made involves reentrant signaling, or correlated activity, between maps that represent information that is received from the different sensory modalities. To illustrate, the visual map would receive inputs from the eyes and represent an image of your grandmother's face. The auditory map would receive inputs from the ears and represent the sound of your grandmother's voice. Because these two maps are always active when you are in the presence of your grandmother, their activities become correlated and a higher-order neural circuit that links them would come into being. This circuit would then conflate the different features of your grandmother into a single coherent representation.

The theory of neuronal group selection is significant because it suggests that many cognitive processes are not hardwired or pre-specified at birth. It asserts that, instead, we are born with a tangled and variable web of neural circuits, which are pruned and selected by experience. Note that this understanding

contrasts with that promoted by EP, which posits the existence of innate mental modules. By now this argument should sound familiar, as it corresponds to another argument—that of the nativist position versus the empiricist position with respect to knowledge, discussed previously in Chapter 2 (The Philosophical Approach). You will recall that nativists believe that we are born possessing knowledge, whereas empiricists think that knowledge is acquired. It is probably fair to say that cognitive abilities are influenced to varying degrees by innate and experiential factors, and that it is the job of future studies to determine the relative contributions of each. Some of the details of the nature-nurture debate are provided in the next section.

Evaluating Evolutionary Psychology

Because we have already alluded to some of the problems that attach to EC, we concern ourselves here with a critique of EP. EP as a formal discipline is a recent addition to the ranks. As such, it has received both fervent acclaim and fervent criticism. Its strengths are its methodological variety and its attempt to develop an objective science of human behavior that is grounded on a single, powerful theoretical model. The evolutionary model has already proved to be fruitful in biology and its use in psychology links the two disciplines. The evolutionary perspective also reshapes psychology such that it becomes a unified discipline, as opposed to the disjointed one that it tends to be currently (Buss, 1999). Modern psychology consists of multiple disciplines, each characterized by a unique theoretical perspective and a unique focus. For example, cognitive psychology focuses on information processing, social psychology focuses on interpersonal relationships, and developmental psychology focuses on growth and change in an organism over its entire lifespan. EP cuts across these different theoretical perspectives and can provide a coherent, unified approach to the study of mind and behavior. But EP has its problems. Even adherents of the theory of natural selection and of biological evolution have attacked its assumptions.

One of the fundamental assumptions behind the evolutionary approach is that evolved psychological mechanisms are adaptive responses to survival-related or reproduction-related problems. But this need not always be the case. There are at least two other kinds of phenomena, in addition to selective processes, that can produce novel biological function (Dover, 2000). The first of these is **exaptation** or **neutral drift**. In this instance, random mutations in an organism's genome result in new genes that have little or no consequence for reproduction. These genes can become disseminated in a population, and at

some point may assume a new adaptive function that represents a response to a change in the environment. The second is **molecular drive**, in which a copy of a gene can mutate. There are known cases in which genes that code for proteins become duplicated; the duplicated genes then diverge to serve another function. These genes and the characteristics they give rise to have not been selected for.

The evolutionary scholar Stephen Jay Gould proposes that many of an organism's attributes may arise via nonselectionist mechanisms. He refers to these attributes as "spandrels" (Gould, 2000). A **spandrel** is the triangular space between the exterior curve of an arch and the rectangular frame that encloses the arch (see Figure 8.12). A spandrel serves no design purpose; it exists solely as a consequence of an architect's surrounding an arch with a frame. Similarly, many psychological mechanisms may themselves be spandrels— they may be byproducts of an adaptation that was subsequently co-opted for useful purposes. The ability to read and write are probably byproducts of an earlier linguistic capacity—the capacity to comprehend and produce speech. Once these earlier capacities were in place, neural connections to visual and manual parts of the brain that were already in existence would then have enabled the later-arriving capacities.

The job of the evolutionary psychologist is a difficult one, as he or she must start with a given mental mechanism and come up with a purpose for it. The purpose corresponds to the adaptive problem the mechanism was intended to solve. This process is called **reverse engineering**, because one begins with the final product and thinks back to what it was that may have compelled its genesis. This is more difficult than the conventional engineer's job, which consists of starting with a desired function (getting a car over a valley) and designing a product (a bridge) that will enable that function. It is also the case that evolutionary psychologists can easily be wrong in making the assumption that a particular psychological mechanism came into being as a response to a particular problem. It may have evolved in response to an entirely different problem.

What makes reverse engineering so difficult is the paucity of available evidence with which one might reconstruct crucial past events. Tools, bones, and other fragments of the archeological record allow us to make certain inferences, but they are insufficient to inform us about crucial aspects of prehistoric times, for example, kinship relations, group size, social structure, and the activities of males and females (Gould, 2000). These aspects of culture and behavior cannot be easily revealed through the fossil record.

In the traditional evolutionary view biology is shaped by selection. Biology then gives rise to psychology. Psychology in turn generates culture—the larger environment in which humans live (Tooby & Cosmides, 1995). This scheme can be turned upside-down, however, in that culture can also shape

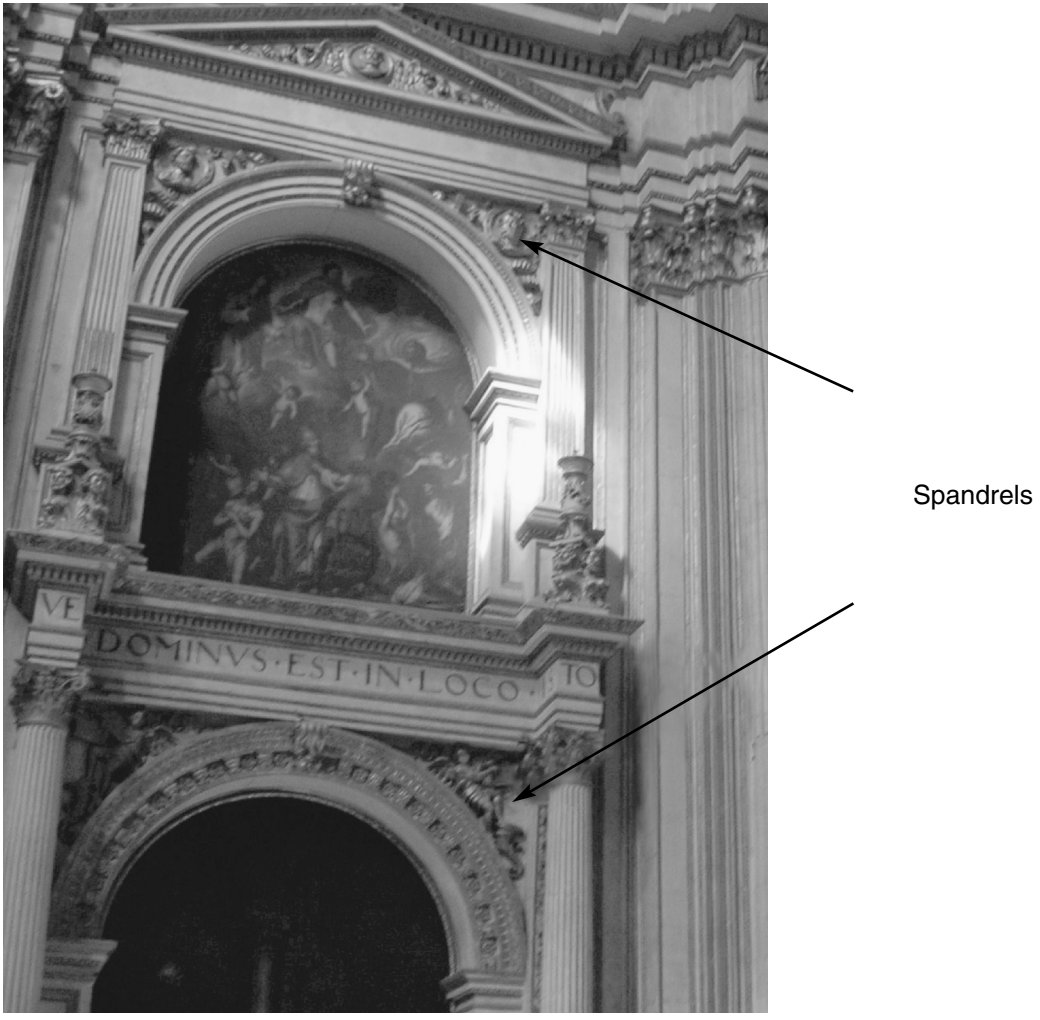


Figure 8.12 Spandrels in the cathedral in Granada, Spain

psychology. A large body of evidence shows that environments exert a strong influence on our thoughts and behaviors. Culture, primarily via the medium of learning, is thus another force that is capable of generating psychology.

Dynamical or **ecological models of cognition**, because they postulate the existence of learning, pose a threat to evolutionary theory (Karmiloff-Smith, 2000). In these models mental abilities arise through learning processes. That is, they come into being as the result of an individual's social, verbal, perceptual, and

manipulative interaction with his or her environment—all taking place within the lifespan of the individual. In this account, what would seem to be an evolved psychological mechanism may in fact have been acquired by the individual as he or she went about his or her business. This runs contrary to the evolutionary view that evolved psychological mechanisms are coded for by genes that were shaped by selection pressures and that were present at birth, fully formed, and just waiting to be activated by the appropriate stimulus inputs.

An alternative account has genes laying down the groundwork for a capacity that develops at a later stage as a result of experience. Stereoscopic vision, the ability to perceive depth via a combination of the information that is received from the two eyes, is an instance of this kind of capacity. Stereoscopic vision is not present at birth, and develops fully only after an organism has perceived the world with both eyes. The brain mechanisms that underlie such abilities cannot exist in their entirety at birth, as developmental experience forges their basic neural connections and circuits—a process termed "post-natal neural assembly."

Karmiloff-Smith (2000) distinguishes between two developmental models. In the **mosaic model of development**, a brain mechanism is determined almost entirely by genes, operates quickly (i.e., is a reflex or fast perceptual process), and its components or parts develop independently of one another. Such mechanisms operate well under typical conditions—those "designed" to trigger their operation. They must, however, be completely specified in advance and are limited in how complex they can be. Evolutionary psychologists believe that evolved brain mechanisms are, in a word, mosaic. In contrast, in the **regulatory model of development** brain mechanisms are only partially determined by genes, operate more slowly (i.e., are deliberative cognitive processes), and their parts develop interdependently. This type of mechanism is flexible and capable of altering its function under varying conditions, and only the broad outlines of such a mechanism need be specified at birth. These mechanisms are also more complex. Research suggests that regulatory development is the norm for the growth of cortical abilities—those that underlie higher-order cognitive abilities. Mosaic development occurs in many animal species, and in humans is likely to govern several non-cortical brain areas.

Overall Evaluation of the Evolutionary Approach

Evolution is a powerful theoretical framework. It can be used to explain how virtually every natural process comes into existence. As such, it rests squarely within Marr's computational level of description, which attempts to provide explanation for a given process by identifying its purpose. Knowing why a

structure exists gives us many clues as to its structure and function and can help us to decipher problems that come into view at other descriptive levels. The evolutionary approach, as we noted earlier, meshes well with neuroscience. Collaboration between these disciplines should focus on finding the neural bases of evolved psychological mechanisms, and on providing a fuller understanding of the selection pressures that created them.

One bone of contention between evolution and the cognitive and artificial intelligence perspectives is the issue of general-purpose vs. domain-specific processing. Is the mind a general-purpose computer or a "Swiss army knife"? As is the case with many other debates that are apt to pit the proponents of one extreme against proponents of the opposite extreme, it may be that both are correct. Some cognitive processes may indeed qualify as evolved psychological mechanisms, by their being in accordance with all of the defining criteria. Others may be general-purpose. There is a need for studies that will focus on this important distinction and on the role of complex developmental factors.

In Depth: A Stage Theory of Evolution

Merlin Donald, in his book *Origins of the Modern Mind* (1991), proposes a detailed three-stage model of how the human mind may have evolved. He believes that primordial humans were at first not very different from other primates such as chimpanzees. Primate cognition is characterized by episodic memory, the representation of information that is concrete or time-bound. Animals with episodic ability are limited to processing information that pertains to their direct experiences—those experiences linked to a specific location and time. An ape can understand and remember that it was groomed by its comrade yesterday, but would have trouble thinking about hypothetical or abstract situations, for example, that it would need to eat a large quantity of food before setting off on a long trip. This would suggest that apes, unlike humans, do not possess semantic memory. Semantic memory, as discussed earlier in the section on memory, corresponds to the ability to represent general world knowledge, which would include facts like "Dogs have four legs."

Episodic ability, though, allows for a wide variety of adaptive skills. Primates are capable of learning through observation and imitation, as this requires only the representation of a specific event and the execution of a response. They also use simple tools. Chimpanzees can take a stick, poke it into an ant or termite nest, and then pull it out and eat the insects that cling to the stick. A number of non-human primates have been taught language-like abilities. Some have been taught American Sign Language, and others have been

able to communicate via pictorial representation. Although primates can learn basic linguistic skills from human instructors, they don't consistently use the skills on their own, teach them to other primates, or generate novel sentences. These issues are discussed at greater length in the chapter on language.

According to Donald, the first stage in the development of the human mind was the emergence of **mimetic skill**, the ability to convey or act out information via voluntary motor actions. Donald defines mimetic skill as the ability to produce self-initiated representational acts that are intentional but not linguistic. These acts are not mimicry or simple copying. Mimetic skill includes facial expression, manual signs and gestures, posture, and whole body movements. It is used to communicate information about an event to others. It thus serves as a precursor to speech. Mimetic skill has many language characteristics, including generativity, the capacity to generate an unlimited number of meanings. Mimetic skill, like language, serves a social function. It allows for communication among individuals in a group. It imparts a tremendous survival advantage, because group members can then engage in cooperative social behavior, for example, planning a hunt and sharing information that is related to the hunt.

In stage two, human societies have undergone transition and become mythic cultures. A myth requires the construction of a conceptual model of the human universe. Myths are present in the history of nearly every major culture. They tell stories about the creation of the universe, heroic adventures, and war. Myths are oral and therefore require spoken language, the ability to speak and listen. This form of storytelling is vocal or phonological and does not rely on an external storage system such as writing. Myths are represented solely in the minds of those who have heard them and are transmitted from generation to generation. A myth integrates multiple events in a temporal and causal framework. Its primary level of representation is thematic, rather than event-based.

Humans capable of myth have taken a large cognitive leap. They possess spoken language ability and can represent events and complex ideas not limited to one place or time. These skills are built upon the strong survival advantage that has been bestowed by mimesis and include improved social coordination, tool manufacture, shelter construction, and food acquisition. Improved social exchange in all likelihood also led to more complex social and political structures and the laying down of rules and regulations having to do with marriage customs, property ownership, and the resolution of conflicts.

The third stage in human cognitive development, says Donald, is embodied by the creation of an external symbolic storage system. The first such systems were pictographic, for example, Babylonian cuneiform and Egyptian hieroglyphics. These systems were succeeded by alphabets. They heralded a new human capacity—that of reading and writing. We see here a shift from

internal to external storage, from information that resides only in the biology of the human brain to its being represented via an external device that entails technology. There has also been a shift from narrative to paradigmatic thought (Bruner, 1986). **Narrative thought** is subjective and holistic. It is employed more often in the arts and humanities. **Paradigmatic thought** is objective and analytical and is implemented more often in the sciences. Donald believes that the development of symbol systems and the formal manipulation of symbols catalyzed the development of theory, objectivity, logic, and reason in humans. These skills are themselves adaptive, as they allow for greater understanding of and technological control of the natural world.

Minds On Exercise: Memory for Object Location

Pair up with a friend of the opposite sex. One of you will then select five objects from Figure 8.11, without telling the other member of the pair what they are. The other member of the pair will then study the figure for one minute. He or she will then describe the locations of the five objects. Now switch places, and repeat the exercise, using a different set of five objects. How did each of you provide an answer? Did males describe the locations of the objects using frame of reference terms such as "above" and "to the right of"? Did females code object position by using landmarks, that is, by referring to other nearby objects? Which sex did better overall?

Food for Thought: Discussion Questions

1. Are all mental capacities evolved psychological mechanisms? Can you think of some that might not be? Consider the "six properties" of evolved psychological mechanisms. Do these properties describe the mental capacities that you've chosen?
2. The psychologist Paul Ekman (1989) proposes five basic human emotions. They are anger, disgust, fear, happiness, and sadness. What do you think were the selection pressures that led to each? What adaptive purposes do these emotions serve? Do they continue to be adaptive in modern society?
3. Clinicians identify anxiety, depression, sociopathy, bipolar disorder, and schizophrenia as major categories of psychological disorder. Do these disorders represent breakdowns in normal mental functioning or were they selected for during the EEA? Could any of these disorders have served an adaptive function?

4. What are the selection pressures, other than those discussed in this chapter, that may have led to the development of human intelligence? Formulate a stage model for the evolution of intelligence.
5. There is evidence to suggest that females are better than males at understanding facial expressions and at responding to other social and emotional cues. What do you think is the evolutionary explanation for this?
6. Women rely more on landmarks when finding their way around a new environment (Sandstrom, Kaufan & Huettel, 1998). Men rely more on frames of reference, such as the directions "north" and "south" or "left" and "right." What do you think is the evolutionary explanation for this difference?

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Buss, D. M. (1999). *Evolutionary psychology: The new science of the mind*. Boston, MA: Allyn & Bacon.

Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind*. New York: Oxford University Press.

Darwin, C. (1859). *On the origin of species*. London: John Murray.

Dawkins, R. (1989). *The selfish gene*. Oxford, UK: Oxford University Press.

Edelman, G. M. (1989) *Neural darwinism: The theory of neuronal group selection*. Oxford, UK: Oxford University Press.

Gaulin, S. J. C., & McBurney, D. H. (2001). *Psychology: An evolutionary approach*. Upper Saddle River, NJ: Prentice-Hall.

Halpern, D. F. (2000). *Sex differences in cognitive abilities*. Mahwah, NJ: Lawrence Erlbaum.

Kahneman, D., & Tversky, A. (Eds.) (1982). *Judgement under uncertainty*. Cambridge, UK: Cambridge University Press.

Levy, S. (1992). *Artificial life*. New York: Pantheon.

Mitchell, M. (1996). *An introduction to genetic algorithms*. Cambridge, MA: MIT Press.

Rose, H., & Rose, S. (Eds.) (2000). *Arguments against evolutionary psychology*. New York: Harmony Books.

9

The Linguistic Approach: Language and Cognitive Science

"How very commonly we hear it remarked that such and such thoughts are beyond the compass of words! I do not believe that any thought, properly so called, is out of the reach of language."

—Edgar Allan Poe, 1846

The Linguistic Approach: The Importance of Language

Linguistics is the study of language. There are many different kinds of linguistics studies—each with its own theoretical perspectives and methodologies. Some of these adopt a neuroscience approach and use the case study method; the researchers study the language-related deficits of patients who have suffered brain damage. Others implement various network models of how language information is represented and processed. Some linguists take on a developmental orientation: they examine how language ability grows and changes with time during the development of the individual. Still others who study linguistics are philosophers who ask questions about the nature of language and of the relationship between language and thought. In fact, language can be studied from the vantage point of every field of study that has been

described in this book and more. What makes linguistics unique, then, is not the perspective or the tools it brings to the table, but the subject matter of the investigation—language itself.

Linguistics studies attempt to answer many questions. Language is so complex that much of the research that has been conducted in this area has been directed toward an understanding of the structure of language (in addition to how it is used). These studies have focused on grammatical rules that specify allowable combinations of linguistic elements. Another interesting issue is whether humans are unique in using language, or whether some animals possess language ability. Languages of course cannot be learned overnight, and many linguists have studied language acquisition—how it is acquired during development. Linguistics is truly interdisciplinary, and in this chapter we survey the approaches to the study of language that are represented by the disciplines of philosophy, cognitive psychology, neuroscience, and artificial intelligence.

The Nature of Language

With this in mind, we begin our discussion of the linguistics approach with an exploration of the nature of language. There has been much debate about what language is, exactly—and there is no agreed-upon definition. It is easier to list its most important characteristics (in lieu of providing a definition). According to Clark and Clark (1977), language has five characteristics:

1. **Communicative.** Language allows for communication between individuals. *Communication* refers to the production, transmission, and comprehension of information.

2. **Arbitrary.** A language consists of a set of symbolic elements. Symbols, as we noted in Chapter 1, are referential, they stand for or refer to something. Linguistic symbols can be almost anything. Most commonly, these symbols are sounds, pictures, or words. The defining hallmark of these symbols is that they are completely arbitrary. Virtually any sound, picture, or word could be chosen to represent a particular thing. The sound of the word that represents "house" in English is different from the sound of the word that represents the same item in Spanish.

3. **Structured.** The ordering of the symbols in a language is not arbitrary, but is governed by a set of rules. The rules specify how the symbols may be combined. In English, we place the adjective before the noun, as in the

phrase: "the big house." In Spanish, this same proposition is expressed via a different set of rules: the adjective follows the noun, as in: "la casa grande."

4. **Generative**. The symbolic elements of a language can be combined to create a very large number of meanings. Just think of how many six-word sentences one would be able to generate in English. If we start with the sample sentence "The fox jumped over the fence," we could then substitute "dog," "cat," "deer," and words for many other animals, for "fox." Likewise, we can substitute "bottle," "can," or "tire" for the word "fence." So the number of variations on just this one sentence is large. Every day, we utter new sentences that we have never uttered before. The generative property of language makes language very powerful, as virtually any idea that can spring to mind can be expressed.

5. **Dynamic**. Languages are constantly changing as new words are added and grammatical rules altered. Only thirty years ago, there was no word for the concept that "e-mail" represents—because it didn't exist.

There is a fundamental distinction to be made that has to do with the type of linguistic representation: whether that representation is auditory, visual, or having to do with another sensory domain. Spoken language is naturally produced via the faculty of speech and is understood via listening. Speech and listening to speech can of course be transformed into their equivalents within the visual domain—writing and reading. Beyond this, there are also languages that consist of motoric gestures, such as American Sign Language (ASL), and tactile languages, such as Braille. If we consider spoken language, the most common form of language usage, we must then describe two important elements of spoken language. These are the phoneme and the morpheme.

A **phoneme** is the smallest unit of sound in the sound system of a language. A phoneme has no meaning. Phonemes correspond in a rough way to the letters of an alphabet; in some instances multiple phonemes correspond to a single letter. The phoneme for the letter "a" as it is pronounced in "father" corresponds to the "ah" sound, whereas the phoneme for "a" as it is pronounced in the word "cane" corresponds to the sound "ay." There are about 45 phonemes in the English language. Some instances are shown in Table 9.1. The smallest number of phonemes reported for a language is 15. The largest is 85. Phonemes, like letters, are combined to form the spoken versions of words.

Morphemes are the smallest units of spoken language that have meaning. They roughly correspond to words, but can also be parts of words. Thus, the sound of the spoken word "apple" is a morpheme, but so is the sound of "s"

Table 9.1 Selected English consonant and vowel phonemes

Consonants	*Vowels*
p (pill)	i (beet)
w (wet)	e (baby)
s (sip)	u (boot)
r (rate)	o (boat)
g (gill)	a (pot)
h (hat)	^ (but)

denoting the plural form. If we want to change the form of "apple" from singular to plural, we add "s" to form "apples," which changes the meaning. Similarly, there is the morpheme that corresponds to the sound "ed," which, when added to the root form of many verbs, forms the past tense. Considering that there are about 600,000 words in the English language, the number of morphemes that the language has is quite large.

In addition to the elements of language, there are the rules that allow for their possible combinations. There are multiple sets of rules. **Phonology** refers to the rules that govern the sound system of a language; **morphology,** to the rules that govern word structure; **syntax,** to those that govern the arrangements of words in sentences; and **semantics,** to those that have to do with word meanings. Collectively, these rules are known as the **grammar** of the language. It is important to distinguish between the meaning of the word as it is used most commonly and its more esoteric meaning (used in linguistics studies). Prescriptive grammar is the formal and proper set of rules for the use of language, in which we all received training at school. Descriptive grammar refers to the underlying rules, which linguistic researchers infer from the way people actually use language. We will come back to the subject of grammar and how it is used to describe the hierarchical structure of sentences in our discussion of Noam Chomsky's theory of language.

Language Use in Primates

Animals in the wild communicate with one another. A monkey species that lives on the African savannah has a specialized series of cries that signify different kinds of threats. The monkeys use these cries while they are feeding to warn each other of impending danger. If one monkey in a group spies an eagle

circling overhead, it emits one type of cry, which sends the members of the groups scattering into the trees for cover. If it spots a snake it emits another cry, which impels the monkeys to stand up on their hind legs and look around so that they can try to locate the snake. Each of the cries has a specific meaning. Each stands for a particular danger to the group. The meaning of the cry is understood by the rest of the group, as indicated by their reactions. This is communication, because information about an event was produced, transmitted, and comprehended. But this natural form of communication is not language. The cries are not arbitrary, there is no use of a grammar to arrange them into anything like sentences, and they are not combined to create new meanings.

This raises an interesting question. If animals don't use language naturally on their own, can we teach it to them? Do they have the same capacity for language that we do? Research in this area has focused on primates, such as chimpanzees and gorillas, because of their relatively advanced cognitive capacities. Let's summarize some of this research and evaluate the results.

Early investigations of the linguistic abilities of primates focused on language production. Animals evidenced rudimentary language skills after being trained in one of several linguistic systems that included ASL, as well as a symbolic system employing plastic tokens and one that used geometric patterns called **lexigrams**. Starting in the 1960s Beatrice and Allen Gardner raised a chimp named Washoe. They taught her to use ASL. Their method was to get Washoe to imitate or reproduce the hand formation that stood for a particular object (Gardner, Gardner & Van Cantfort, 1989). Washoe learned 132 signs and seemed to show evidence of spontaneous language use. Upon seeing a toothbrush in the bathroom, she made the sign for it without being prompted. A similar technique was used to teach ASL to a gorilla named Koko (Patterson, 1978). Koko learned an even larger repertoire of signs and was reported to have used syntax and to have made signs spontaneously. Her trainer claims she even told jokes!

David Premack has used a different approach. He used plastic tokens instead of hand signals as he attempted to teach language skills to a chimp named Sarah (Premack, 1976). The tokens had different shapes and colors and stood for individual words as well as relationships. There were tokens that stood for nouns ("apple"), for verbs ("give"), for adjectives ("red"), and for relationships ("same as"). Sarah produced the "same as" token when she was presented with two apple tokens, and the "different" token when shown an apple and an orange token. She seemed to have a rudimentary understanding of sentence grammar, as she was apparently able to tell the difference between two sentences such as "David give apple Sarah" and "Sarah give apple David."

Savage-Rumbaugh et al. (1993) studied a chimp named Kanzi who appeared to have learned the meanings of lexigrams. In addition, Kanzi was apparently able to understand single-word and simple-sentence utterances made by humans. Kanzi's abilities seemed quite advanced. Of his own accord, he would use lexigrams to identify objects, to make requests for food items, and to announce a particular action he was about to undertake. Following more structured language training, Kanzi's abilities were compared to those of a two-and-a-half-year-old child named Alia. Both were given novel commands that required them to move objects. In terms of comprehension, the two showed nearly identical abilities: they both demonstrated approximately 70% compliance with the commands. Kanzi's language production skills were more limited—they corresponded only to those of a one-and-a-half-year old child (Greenfield & Savage-Rumbaugh, 1993).

Evaluating Language Use in Primates

At this point we can examine some of the criticisms that have been leveled at this research. Some of the animals described so far were trained via the use of positive reinforcement. They were given a reward, usually a food item, for making the correct sign or using the appropriate token or lexigram. A problem with this is that the animals may have been associating a symbol with a concept because they had been trained to do so, and may not have had any awareness that the symbol actually stood for something. If this were the case, these animals fail to demonstrate the arbitrariness aspect of language—that the symbol can be anything and still stand for its referent. An animal's choosing an apple token when presented with an apple does not demonstrate arbitrariness. But using an apple token to refer to an apple when an actual apple is not perceptually present does. This aspect of language, in which users refer to something that is removed in space or time, is called **displacement**.

Savage-Rumbaugh (1986) presents some evidence of displacement in chimps. She employed a technique known as cross-modal matching wherein chimps who viewed a lexigram were then required to select the object the lexigram represents via the use of touch from a box that was filled with objects. The chimps were able to do this, which indicated they understood what the lexigrams represented. A note of caution is in order, however. Arbitrariness and displacement capabilities were shown for a comprehension task only, where the animals were interpreting the symbols. Earlier studies in which chimps used ASL and tokens have generated less evidence that primates understand the meanings of symbols when they produce them.

What about the structured aspect of language? Do animals understand the syntax that underlies the formation of sentences? The investigations that have been conducted thus far show that primates comprehend and produce very simple sentences—sentences that are on the order of two or three words long. An understanding of the rules of syntax is demonstrated by the ability to rearrange words in new combinations that express new meanings—the generative criterion, defined above. If animals could do this, it would indicate a comprehension of syntactical rules.

The researcher Herb Terrace provides us with evidence that refutes the idea that some animals may have a rudimentary understanding of or the ability to use syntax (Terrace et. al., 1979). Terrace was skeptical that chimpanzees like Washoe truly understood the meanings of the signs and symbols they used. As alluded to above, he believed that chimps used hand signals or presented tokens because they had been reinforced for doing so. To test the idea, he studied a chimpanzee whom he had jokingly named Nim Chimpsky. Nim was raised in a human family and taught ASL. Rather than use food as reward, Terrace gave approval that centered on things that were important to Nim. Under this system Nim did seem to have some grasp of the meanings of his signs, as he was found using them in the absence of their referents. He also appeared to use signs spontaneously to express his desires. For example, he would make the sign for "sleep" when he was bored. However, Terrace concludes that Nim was never able to combine his signs so as to form sentences and express novel meanings. He did this only when he was directly imitating combinations of signs that had been produced by his trainers. Figure 9.1 shows Nim Chimpsky learning a sign.

So at this point we can sum up the work on the language abilities of primates. Primates appear to possess some arbitrariness and displacement capabilities because they can comprehend the meanings of a limited number of symbols independent of their referents. This is true whether they are trained directly (with food) or indirectly (with approval) using positive reinforcement techniques. But here is where their linguistic skills seem to come to an end, for primates seem to understand very little in the way of syntax, especially when it comes to language production. They know that some aspects of word order affect meaning. However, most primates tend to repeat the sentences they were taught or produce only small variations on them. They do not come anywhere near to possessing human generative capability. Also, unlike humans, primates, once they have acquired language skills, fail to teach the skills to other members of their species. Unfortunately, the "Doctor Doolittle" scenario of our conversing with animals the way we do with one another just doesn't seem possible.

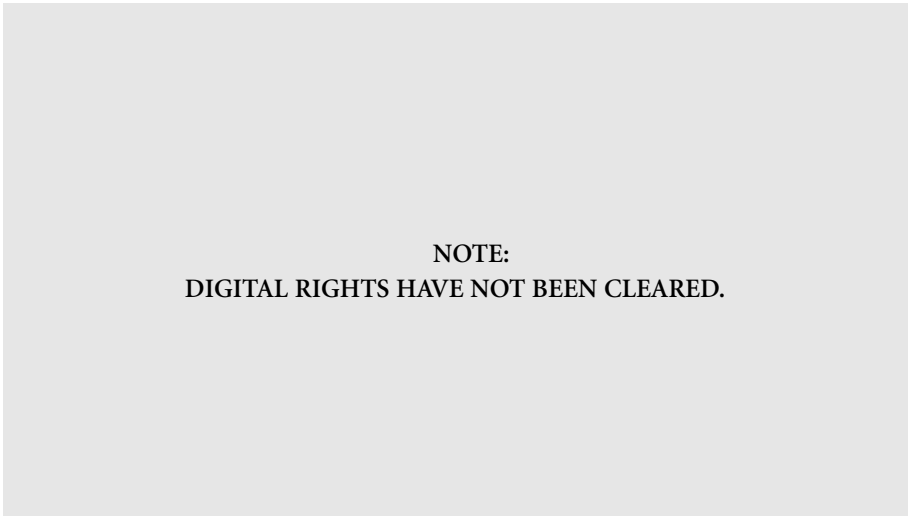


Figure 9.1 Nim Chimpsky signs "black."

Source: Photo courtesy of AnthroPhoto

Language Acquisition

Clearly, a human being is not born with an ability to speak his or her native language fluently. This ability develops over time. Linguists adopting a developmental perspective have studied the acquisition and development of language skills, from birth through infancy, childhood, and into adolescence. They have shown that human beings pass through a series of stages, each one marked by the acquisition of new linguistic skills.

Early in the first year, infants start to utter a wide variety of sounds. All infants do this. At this stage they begin to exercise their vocal cords and mouths—major parts of the vocal apparatus, the use of which they must master in order to articulate the sounds of the language they are just beginning to acquire. This period of development is known as **the cooing stage**. Figure 9.2 shows an infant communicating with her caregiver.

At around six months, the number of sounds a baby produces has shrunk. The sounds produced at this stage are consonant-vowel pairs, such as "mama" and "dada." The majority of the utterances made at this time are more phonemic than morphemic in nature. They correspond to sound units rather than fully pronounced words. However, the intonations of these utterances at this point begin to match those of the language the child is learning. Intonation

Figure 9.2 Early in development, infants will spontaneously babble phonemes

refers to the rises and falls in pitch and changes in other acoustic properties of one's speech. For example, for most speakers there is usually a rise in pitch at the end of a question. These abilities arise during the so-called **babbling stage**.

Following the babbling stage and just a few months shy of the child's first birthday, we see the advent of one-word utterances. At this point children are able to successfully articulate entire morphemes. These morphemes, or words, may not be prescriptively accurate. A child may say "unky" instead of "uncle," but the utterance is being used in a meaningful way. The children are thus beginning to use language in a symbolic and semantic fashion. This is the **one-word stage**.

Following this, during the **two-word stage**, children produce two-word utterances. It is at this point that they will say things like "see kitty" or "want toy." Because words are now being arranged into simple sentence-like structures, the two-word stage marks the emergence of rudimentary syntactical skills. After the two-word stage, babies will string together more complicated utterances, speaking out the equivalent of sentences composed of three or more words that convey increasingly complex meanings.

There are no clearly identifiable stages that follow the two-word stage. But this period is characterized by a steady growth in vocabulary and syntax. Also during this period, children exhibit some interesting patterns of development, especially in regard to their learning of past tense forms (Kuczaj, 1978; Marcus et al., 1992). Studies of this type of learning show that children first imitate past tense forms correctly. For the irregular verb "to go," the past tense form is "went." After children have learned the general rule of past tense formation, they apply it correctly to regular verbs, but then overextend the rule to include irregular verbs as well, saying, for example, "goed" instead of "went." Finally, they learn the exceptions to the rule, for example, using "went" only when it is appropriate. This intriguing U-shaped pattern of development indicates the presence of several learning strategies in children: they start out with purely imitative copying, proceed to an understanding of a rule, and ultimately progress to the learning of the exceptions to that rule.

Evaluating Language Acquisition

One of the first questions that springs to mind in discussions of language acquisition is whether language ability is innate or learned. This is of course one of the nature-nurture debates. We have previously addressed this question in regard to language. In the introduction to Chapter 4 (The Cognitive Approach I) we said that the fall of behaviorism was due in part to its inability to account for an environmental theory of language acquisition. Researchers found that children learn words and syntax far more quickly than could be explained by imitation or reinforcement. Also, in the section on language in Chapter 8 (The Evolutionary Approach), we presented evidence that language is an evolved capacity. This evidence includes the universality of some characteristics of language, the linguistic stages that all babies progress through, and the existence of specialized brain areas devoted to language processing. This preponderance of evidence indicates that many aspects of language are genetically specified. It does not mean that the environment plays no role in language acquisition, only that we must work to uncover the interplay between environmental conditions and the operation of any innate mechanisms.

Language Deprivation

What, then, is the role of environment in language acquisition? One very basic approach to this question is to examine language ability in the absence of exposure to language. If experience of language is necessary and innate linguistic

mechanisms are dependent on it, we should see language deficits in the absence of exposure to a normal linguistic environment. If, on the other hand, experience and stimulation have little to do with language development, then language ability should remain relatively intact in their absence. Studies that have investigated this issue have demonstrated the existence of a pivotal time in development during which language must be learned. Children not exposed to language during this time, called the **critical period**, may never acquire it or may suffer severe language impairments.

There is abundant evidence in support of the idea of a critical period. Let's examine some of it. The first evidence comes from studies of birds. It turns out that some birds (as well as human beings) are among the few animals that need to be exposed to the communicative sounds of their own species in order to be able to produce them (Doupe & Kuhl, 1999). Both birds and human beings are better able to acquire communicative ability early in life. Marler (1970) showed that white-crowned sparrows after the age of 100 to about 150 days were unable to learn new songs by listening to a "tutor" bird. This was true for birds that were raised with such tutors or were exposed to taped examples of bird song, as well as those that were raised in acoustic isolation.

Another line of evidence that supports the existence of a critical period comes from studies of the acquisition of second languages. So far, we have been discussing first-language acquisition, wherein a single language is learned. It is often the case of course that people learn to speak more than one language. Researchers can study the difficulty with which an individual acquires a second language in relation to the time of onset of the exposure to the second language. One such study found that native speakers of Chinese and Korean (for whom English was a second language) received scores on tests of English grammar that bore a relation to the time of onset of their exposure to English: the later their age at time of arrival in the United States, the lower were their scores (Johnson & Newport, 1989). Figure 9.3 shows the results of this study.

The most emotionally compelling evidence that supports the existence of a critical period consists of individual case studies of persons who were deprived of language experience during early development. These cases are tragic but provide a unique opportunity to examine the effects of this kind of deprivation in humans. One famous historical case study is that of the wild boy of Aveyron, who was discovered in a French forest in 1797 (Lane, 1976). The boy, named Victor, had apparently lived much of his childhood completely alone and had very little language ability. He came under the supervision of a physician, Dr. Jean-Marc-Gaspard Itard (1775–1838), who studied him intensively and attempted to teach him language. Despite Dr. Itard's best efforts, Victor never acquired more than the most basic comprehension and production skills.

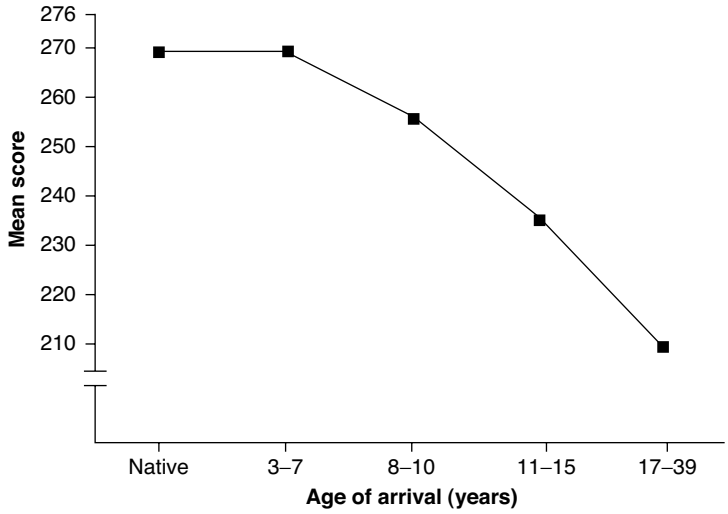


Figure 9.3 Mean English grammar test scores drop the greater the age of a child who speaks another language at time of arrival in the United States

A more recent case study is that of a girl named Genie. Genie spent much of her early life in social isolation. The evidence suggests that the period of her deprivation began when she was 20 months old and lasted until she was "discovered," at the age of 13 years and 9 months. During this time, Genie was kept in a small room, where she was tied to a potty chair for much of the day and night or confined to an infant crib. The door to the room was almost always closed and the windows were closed and covered with curtains. Except for her quick feedings, Genie received almost no parental care. She was not spoken to and there was no radio or TV in the household. Thus she was exposed to little or no spoken language. Furthermore, Genie was beaten for making any sounds of her own.

Given this extreme and prolonged lack of exposure to any sort of linguistic stimulation, what were Genie's abilities like? A number of researchers have chronicled her progress (Fromkin et al., 1974; Jones, 1995). Upon initial examination, she was found not to vocalize at all. Within a few days, she began to respond to the speech of others and to imitate single words. Within a year or so, she was able to understand and produce some words and names. Despite these modest gains, it was clear at the end of the testing period that Genie possessed minimal grammatical ability. Following a period of several years' worth of further evaluation and training, Genie did show signs of simple grammatical

comprehension. For example, she was able to distinguish between the singular and plural forms of nouns, and between negative and affirmative sentences.

At eight months after discovery, Genie uttered two-word phrases such as "yellow car." Later, she was able to produce three- and four-word strings, such as "Tori chew glove" and "Big elephant long trunk." She also demonstrated generativity—she was able to express new meanings by combining words in novel ways.

These capacities show that Genie was able to acquire language in the aftermath of the deprivation period and that the stages of her development, such as her use of progressively longer sentences, paralleled language acquisition in nondeprived children. However, Genie's abilities deviate from those of control children in several ways. Her grammatical ability at the time of early testing was equal to that of a two-and-a-half-year-old child and her speech production capacity was limited. In addition, Genie's rate of language development was slowed in comparison to that of controls. She had difficulty using language to express questions, and many of the hallmarks of language mastery in adults, such as the use of demonstratives, particles, rejoinders, and transformation rules, were absent (Fromkin et al., 1974). In summary, Genie shows that language acquisition following extended deprivation is possible, but that it is severely impaired. To date, Genie has not, and probably will not, develop complete adult language skill.

Evaluating Language Deprivation

Case studies of language-deprived children yield a wealth of information about the individuals under study. They do, however, suffer from a number of problems. To begin with, it is difficult to make generalizations from evidence acquired from a single person or small number of persons. The findings of case studies do not necessarily generalize to a larger population. Second, the conditions that shaped these subjects are often unknown. In the case of Victor, we do not know the duration of his social isolation, or even if he was isolated at all. Some have speculated that he may simply have had a learning disability or had suffered brain damage. With regard to Genie, it is not clear exactly what kind of language information she was exposed to during her formative years, nor the extent to which she may have vocalized to herself.

Research that has investigated the critical period shows that, although there may be an innate language-learning mechanism, it is dependent on environmental input for its proper functioning. If this input is absent, the ability to use language fully never appears. Exposure to and practice in the use of a language is thus a component essential to the development of language. This is true

regardless of the amount of or sophistication of the neural machinery dedicated to language processing in an individual that may be in place from birth.

Philosophy and Linguistics: The Linguistic Relativity Hypothesis

A principal focus of this book is the nature of thought. If thought is representational, as it most surely seems to be, then what is its form? In Chapter 1, we described several forms that thought could take. These included images, propositions, and analogies. Assuming we can think in all of these different formats, then thoughts may assume multiple forms—thoughts may be sometimes pictures, sometimes propositions, sometimes other symbolic representations.

But if we were to vote for the form that we believed thoughts are in most of the time, language would probably win. When we think, it is as if we can hear ourselves talking—what is called implicit speech. In contrast to the imagery that occupies "the mind's eye," implicit speech seems to occupy "the mind's ear." This supposed mental primacy of language has led some to conclude that thought and language are so similar that it may be impossible to express the thoughts generated in one language in another language. This is the strong version of the **linguistic relativity hypothesis**, which also goes by the name of the Sapir-Whorf hypothesis, after the linguist Edward Sapir (1884–1939) and his student, Benjamin Lee Whorf (Whorf, 1956). The weak version denies that such translation is impossible, but admits that the language a person speaks influences the way he or she thinks.

Whorf studied the Hopi language, a Native American language, and found that the Hopi experience time as a discrete series, with each unit of time, say days, considered unique and different from the others. This differs from the Western conception of time, wherein time is experienced as an undifferentiated continuous flow. Thus, a Hopi individual would not say "I stayed five days," but "I left on the fifth day" (Carroll, 1956, p. 216). The strong version of the linguistic relativity hypothesis would argue that the Hopi are incapable of thinking of time as continuous because they lack the words to express the concept in their language. The weak version would argue that the Hopi can understand this concept of time, but that such an understanding would require a re-expression of it that used a completely different set of Hopi words.

So which version of the linguistic relativity hypothesis is correct? Investigation of the hypothesis has proceeded along two avenues. The first has focused on color perception, the second on counterfactual reasoning.

Davies and Corbett (1997) conducted research that tested English, Russian, and Setswana speakers. Setswana is a language spoken in Botswana, Africa.

Davies and Corbett gave the participants colored chips and asked them to arrange the chips into groups based on chip similarity in any way they wanted. If the strong version of the hypothesis were correct, the Russians would place light and dark blue chips in separate groups, as their language has distinct color terms for light and dark blue. The Setswana speakers would group green and blue chips together, because they have a single term for green and blue. Their study and a follow-up study showed that all participants tended to group the chips in pretty much the same way, regardless of their linguistic background (Davies, Sowden, Jerrett, Jerrett & Corbett, 1998). The studies, therefore, fail to support the strong version of the hypothesis.

A counterfactual statement is a hypothetical one. It asks us to imagine what would happen if something were true. Counterfactuals can be expressed in two ways. One way is through the use of the subjunctive mood. The statement "If you bought your ticket now, you would save money" is an example of a statement that uses the subjunctive mood. The use of the word "would" is part of the subjunctive application and signals that it is an imaginary scenario that is being expressed. Other forms of the subjunctive use "were to" or "ought." The other way to express counterfactual statements is through the use of "if-then" constructions. "If it is hot today, then I will go swimming" is an instance of this construction.

Bloom (1981) set out to test the linguistic relativity hypothesis by taking advantage of the fact that English speakers and Chinese speakers have different ways of expressing the counterfactual (see Figure 9.4). English has both subjunctive mood forms and if-then constructions, whereas Chinese has only if-then constructions. He predicted that Chinese speakers would therefore have greater difficult in understanding the counterfactual. He presented both groups of speakers with a hypothetical scenario and found that Chinese speakers failed to grasp the nature of the scenario, while the English speakers did not have difficulty. The results of his study support the hypothesis.

Both the color naming and counterfactual reasoning studies have been criticized on the grounds that tests for either color naming or counterfactual reasoning represent bad ways of testing the linguistic relativity hypothesis. Color perception is in large part the result of the physiology of the visual system, which is the same in everybody. This means that the way we see determines the names we have for colors—not the other way around. Consistency in color naming across populations and cultures supports this idea (Berlin & Kay, 1969). Bloom's original study has also been criticized on the grounds that the scenario it employed was not translated well into Chinese (Au, 1983; 1984). When better translations were provided, Chinese speakers' comprehension of the scenario improved dramatically.

Figure 9.4 Does speaking Chinese make you think differently?

Evaluating the Linguistic Relativity Hypothesis

In summary, the two avenues of investigation fail to provide emphatic support for the strong version of the linguistic relativity hypothesis. However, more recent investigations of the mental representations of numbers (Miura et al., 1993) and the use of classifier words such as "this" and "that" (Zhang & Schmitt, 1998) do provide limited support for the weak version of the hypothesis. More research is needed to establish a definitive answer to this issue. Currently, most investigators believe that languages are powerful enough and flexible enough to express any number of ideas. We can conclude that languages influence, but don't necessarily determine, the way we think.

We should keep in mind that language is but one way of thinking. As mentioned above, there are other forms of mental representation that are not linguistic and that are not governed by linguistic syntactical rules. The formation and processing of visual images seems to constitute an entirely non-linguistic code for thinking. Mathematical thinking and the mental representation of and

computation of numerical quantities, although language-like, may not rely on language mechanisms to operate and could constitute another distinct format for thought. The same can be said for the mental processing of music. Just because language is powerful and flexible doesn't mean it holds a monopoly on thought. If one idea cannot be expressed in terms of another linguistically, this might be achieved via the use of one of these other formats.

Cognition and Linguistics: The Role of Grammar

We said earlier that grammar is a set of rules that governs how words can be arranged in sentences. Grammar is important because it tells us what is a proper way of expressing something in a language, and what is not. If there were no rules or constraints on expression, we could string words together in practically any order and it would be impossible to convey anything. Let's delve a little further into grammar—how it puts constraints on what can be said and how it illuminates several interesting cognitive principles.

Sentences have distinct parts that are hierarchically related. This organization is called a **phrase structure** and can be illustrated via the use of tree diagrams. Figure 9.5 is a tree diagram for the sentence "The big dog chased the black cat." At the highest level, the entire sentence (S) is represented. Moving down one level, the sentence is composed of two parts, a noun phrase (NP) and a verb phrase (VP). Moving down another level, we see that the noun phrase is made up of a determiner (D), an adjective (A), and a noun (N). Determiners are words like "a" or "the." The verb phrase is made up of a verb (V) and another noun phrase that, itself, contains another determiner, adjective, and noun.

There is a grammar that governs the use of phrase structures. This **phrase structure grammar** imposes certain limitations on how a legitimate sentence can be put together. One phrase structure rule is that all sentences are composed of a noun phrase and verb phrase. A second rule is that noun phrases consist of a determiner followed by a noun. Verb phrases can be expressed a bit more flexibly, as a verb followed by a noun phrase, another sentence, or other elements.

A phrase structure grammar is useful for understanding the organization of sentences, but it doesn't tell us how we can rearrange a sentence to express new meanings. Noam Chomsky (1957) was the first to point this out. He notes that a given sentence can be changed in three ways. First, we can turn an active sentence into a passive one, as when "The man read the book" becomes "The book was read by the man." Second, we can turn a positive statement into a negative one, by modifying this original sentence to become "The man did not

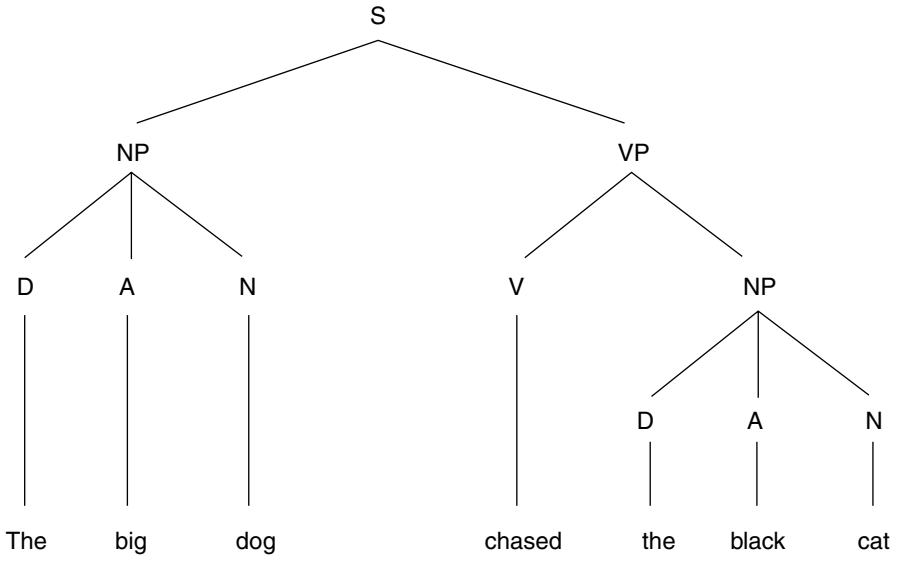


Figure 9.5 The phrase structure for a simple sentence

read the book." Third, we can convert the assertion into a question, as in "Did the man read the book?"

In order to account for these changes, we need a new grammar that allows us to transform one sentence into another. Chomsky's solution was a **transformational grammar**, a set of rules for modifying a sentence into a closely related one. By using these rules, we can reorder "The man read the book" into "The man did not read the book," as follows:

$$\text{NP1} + \text{Verb} + \text{NP2} \rightarrow \text{NP1} + \text{did not} + \text{Verb} + \text{NP2}$$

where NP1 is "The man," the verb is "read," and NP2 is "the book." Similarly, the conversion of "The man read the book" to "The book was read by the man" is denoted as:

$$\text{NP1} + \text{Verb} + \text{NP2} \rightarrow \text{NP2} + \text{was} + \text{Verb} + \text{by} + \text{NP1}$$

An important aspect of a transformational grammar is that one can use it to express two sentences that have different phrase structures but identical meanings. "The man read the book" and "The book was read by the man," above,

have different hierarchical organizations, but they have the same semantic content. To account for this Chomsky proposed two levels of analysis for sentences. The **surface structure** is the organization of the sentence in the form that it is expressed—how the sentence would be heard if it were spoken or read if it were written. The surface structure is variable and can be rearranged by transformational grammar. The **deep structure** is the underlying meaning of a sentence and remains constant regardless of the specific form in which it is expressed.

You may have been wondering whether our discussion so far applies only to English or whether it applies to other languages as well. It is true that languages have different specific rules, but cross-cultural linguistic analyses have shown that languages have a number of elements in common. These commonalities are summed up in the concept of a **universal grammar**, which comprises the features that are instantiated in the grammars of all natural languages (Chomsky, 1986; Cook, 1988). In this view, each individual language at a fundamental level is not really different from others, but represents merely a variation on a theme. Universal grammar is considered as a collection of language rules, hardwired into our brain from birth. In this sense, it is a modular aspect of mind and has all the characteristics of a mental module. It is innate, genetically pre-specified, domain-specific, and independent of other cognitive capacities.

What are the universal properties of all languages then? One is a phonological rule that specifies the ordering of syllables in a word. According to the **maximal onset principle**, consonants usually precede vowels: more frequently than not, they constitute the onset of syllabic groupings. This feature is found in all languages. Another universal is syntactical and concerns the ordering of the subject and object in sentences. In 98% of the world's languages, the subject precedes the object (Crystal, 1987). Thus we say "John kicked the ball," not "A ball John kicked"—even though the latter form is technically acceptable in English.

Universal grammar may be what is responsible for our ability to acquire language so quickly. Language acquisition requires the mastery of a large number of grammatical rules at different levels. There are actually sets of rules, that include phonology, to determine acceptable combinations of phonemes; morphology, to determine which morphemes go together; syntax, for the ordering of words in sentences; transformation rules, for changing the forms of sentences; and so on. The ease and rapidity with which this process occurs in humans can be explained if it is true that at least some generic versions of these rules are already present in the head at birth. A child would then adapt these general linguistic rules to the particularities of the specific language he or she grows up in (Bloom, 1994).

Evaluating Universal Grammar

The idea of a universal grammar or "language organ" as originally formulated by Chomsky has not gone unchallenged. To begin with, there is little evidence to support the notion of specific genes for language. If one looks at other body organs, there are few that owe their existence to individual genes. So it is unlikely that there are specific genes devoted to language processing. There is also doubt about the domain specificity of any proposed language module. The rules governing language use may be more general; they may manifest themselves in other non-linguistic cognitive capacities. One possibility is that linguistic universals are just the product of general biological mechanisms, implying that language is not "special" in any sense.

Neuroscience and Linguistics: The Wernicke-Geschwind Model

Paul Broca (1824–1880) was a French surgeon who worked with patients who had suffered brain damage as a result of stroke or injury. The patients demonstrated various kinds of language deficits, called **aphasias**. Several of his patients had severe difficulty in articulating speech. One famous patient was capable only of uttering the word "tan" over and over. For the most part, these patients could understand what was being said to them, indicating that the faculty of comprehension was intact, but they had problems in pronouncing and producing speech. This deficit is called **Broca's aphasia**. It is also known as nonfluent aphasia.

Patients with Broca's aphasia produce what is called "agrammatic speech." They generate strings of nouns and some verbs, but without any of the filler words such as "the" or "is." They also fail to make words plural or to use verb tenses. Their sentences are short and broken up by many pauses, which has sometimes earned this kind of speech the nickname "telegraphic" or "nonfluent" speech. The following is an example of the speech of a patient talking about a visit to the hospital for dental surgery:

> Yes . . . ah . . . Monday er . . . Dad and Peter H . . . , and Dad . . . er . . . hospital . . . and ah . . . Wednesday . . . Wednesday, nine o'clock . . . and oh . . . Thursday . . . ten o'clock, ah doctors . . . two . . . an' doctors . . . and er . . . teeth . . . yah (Goodglass & Geschwind, 1976, p. 408)

Post mortem examination of the brains of patients who suffered from Broca's aphasia has revealed damage to the lower portion of the left frontal lobe (see Figure 9.6). This region is believed to be at least partly responsible for language production capacity and has been named **Broca's area**.

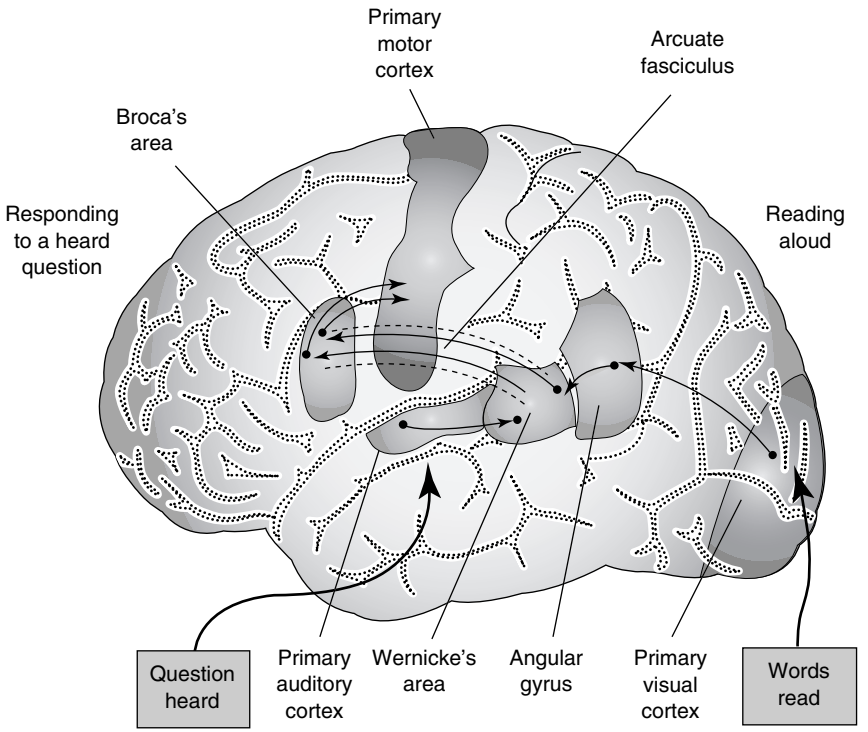


Figure 9.6 Brain areas of the left hemisphere that are part of the Wernicke-Geschwind model of language comprehension and production

A second area, named after Carl Wernicke (1848–1905), mediates language comprehension. This area is located in a posterior region of the left hemisphere (see Figure 9.6). Patients with damage to **Wernicke's area** suffer from **Wernicke's aphasia**. They produce rapid, fluent, and seemingly automatic speech that has little meaningful content. For this reason, this aphasia is also referred to as fluent aphasia. This type of speech sounds normal in the sense that its rate, intonations, and stresses are correct, but it is lacking in content or meaning. These patients have major problems with comprehending speech and also demonstrate difficulty in reading and writing. Here is an example of the speech of a patient with Wernicke's aphasia:

> Oh sure, go ahead, any old think you want. If I could I would. Oh I'm taking the word the wrong way to say, all of the barbers here whenever they stop you it's going around and around, if you know what I mean, that is

> tying and tying for repucer, repuceration, well, we were trying the best that we could while another time it was with the beds over there the same thing. (Gardner, 1974)

The Wernicke-Geschwind model was first formulated by Wernicke and expanded in the 1960s by Norman Geschwind (Geschwind, 1972). It specifies the functional roles of the different brain areas that are involved in language processing, as well as their connections and interactions. Because an understanding of the model relies heavily on an understanding of different cortical areas, we must first introduce these anatomical regions. You may wish to refer back to the neuroscience chapter at this point for a refresher on basic brain anatomy.

Figure 9.6 shows these cortical areas that play the key roles in language processing—as described in the model. For starters, there is the primary motor cortex, located in the frontal lobes in the anterior part of the brain. Commands that originate here send impulses to muscles, causing them to contract and therefore initiating movement; this includes the muscles of the mouth, which must be moved as part of the operation of speaking. The primary visual cortex is located at the back of the brain in the occipital region. It is where visual information is first processed. This area becomes active during reading and writing. The primary auditory cortex is situated in the temporal lobes. It is here where sounds striking the ears are first processed. The **arcuate fasciculus** is a pathway that connects Broca's area and Wernicke's area. Damage to this part of the brain results in an individual's difficulty in repeating words that he or she has just heard, known as **conduction aphasia.** Finally, there is the **angular gyrus**, located behind Wernicke's area. Damage to this part of the brain produces **alexia**, an inability to read, and **agraphia**, an inability to write.

According to the model, these areas and the pathways that connect them subsume language comprehension and production with respect to both the auditory and visual modalities. There is an activation of neural pathways that is the basis of listening and speaking, as follows: The perceptual characteristics of speech sounds would first be processed in the primary auditory cortex. The output of this processing would then be passed to Wernicke's area, where the content of what has been said is processed and understanding is born. A reply is then initiated. From here, the information that will become the reply is passed along the arcuate fasciculus to Broca's area. The information is converted into a motor code, or program of articulation, within Broca's area. This code is then passed to the primary motor cortex, where commands to move the muscles of the mouth and produce the speech that constitutes the reply are executed.

A second pathway meditates reading and writing. In this pathway the primary visual cortex processes inputs that have originated from the words on a printed page. This information is then output to the angular gyrus. The visual

representation of what has been read is converted into an auditory code within the angular gyrus, which then sends the code to Wernicke's area. The remainder of this pathway, responsible for producing behaviors such as reading out loud or writing, coincides with the final portion of the pathway described in the preceding paragraph. The information flow would be from Wernicke's area via the arcuate fasciculus to Broca's area, and then to the primary motor cortex, where muscular action is initiated.

Evaluating the Wernicke-Geschwind Model

The Wernicke-Geschwind model has been criticized on a number of counts. It is considered by some to be an oversimplification of the neural basis for language. To begin with, the areas specified by the model are not completely associated with their hypothesized function. Although in most patients damage to Broca's area or Wernicke's area results in the corresponding aphasias, this is not always the case. Lesions to Broca's area alone produce a transitory aphasia—one that presents with only mild symptoms several weeks after the event that precipitated the injury (Mohr, 1976). More troublesome to the theory are the records of patients with Broca's aphasia who have not sustained damage to Broca's area (Dronkers et al., 1992). The same is true for patients with Wernicke's aphasia (Dronkers, Redfern & Ludy, 1995).

The areas that are specified by the model are characterized as being modality-specific, with Broca's area being a motor-only area that codes for speech articulation, and Wernicke's area being an auditory, sensory-only area devoted to speech comprehension. However, brain-imaging techniques show that these regions are the sites of processing activities that underlie sign-language use (Bavelier et al., 1998). This suggests that they may represent more abstract, modality-independent language ability. In other words, these areas may contain knowledge of syntax that can be applied to any language system, regardless of the modalities involved.

Another criticism of the model centers on its original assumption that these areas are devoted exclusively to linguistic processing. Swinney et al. (1996) have found that lesions in aphasic patients, even those suffering "classic" syndromes such as Broca's aphasia and Wernicke's aphasia, may have disrupted basic processing resources used by the language system. If this were true, damage to the brain areas thought to subsume language only may also lie behind disruptions of other systems that language depends on, such as memory and attention.

Another problem with the Wernicke-Geschwind model is methodological. It was based largely on evidence obtained from clinical case studies of brain-damaged patients assembled after their deaths. Modern science relies more on

brain imaging in live patients. The use of brain-imaging techniques has shown that there are many other brain areas that contribute to language function. We can list a few of them here. The insula lies beneath the frontal, temporal, and parietal lobes. Most patients with Broca's aphasia also have lesions in the insula (Vanier & Caplan, 1990). The left inferior prefrontal cortex, just anterior to and ventral to Broca's area, is activated during semantic retrieval (Peterson et al. 1988). The basal temporal areas, at the bottom of the left temporal lobe, and the cingulate gyrus are also involved in word retrieval. The anterior superior temporal gyrus, anterior to the primary auditory cortex, is implicated in sentence comprehension. These areas are just beginning to be understood. They are believed to interact as parts of a complex network. There is as yet no overarching theory that can describe this interaction. Until then, the Wernicke-Geschwind model provides a useful, if somewhat outdated, understanding of what goes on in the brain during language processing.

Artificial Intelligence and Linguistics: Natural Language Processing

Natural languages are those that have evolved in human societies and are used by human beings, such as English, Spanish, and French. These are in contrast to formal computer languages such as C++, or linguistic expressions of logic. There are two kinds of natural language processing. Understanding a natural language involves an individual's assimilation of linguistic expression in some form, such as speech or writing, extracting its meaning, and then undertaking some action that constitutes a response to this meaning. Understanding is what a computer would need to do if it were to interpret a spoken human command and act on it. Generation is the reverse of this process. It involves taking a formal symbolic representation of an idea and converting it to an expression in English or some other natural language. For example, the idea "It is a sunny day" may initially be stored in a particular format in a computer. A computer would be generating language if it could transform this idea into a spoken utterance that a human being could understand. These two processes are thus the computer equivalent of natural language comprehension and production. In this section, we will concern ourselves exclusively with natural language understanding, as that is the area in which research has been concentrated.

Cawsey (1998) outlines four stages of natural language understanding. We will preview each of them, in the order that they occur:

1. **Speech recognition** is the first step in the process, whereby the acoustic speech signal is analyzed to determine the sequence of spoken words.
2. In **syntactic analysis** the word sequence is analyzed via the use of knowledge of the language's grammar. This yields the sentence structure.
3. Following this, the sentence structure and the meanings of the words are used to derive a partial representation of the meaning of a sentence. This is the **semantic analysis** stage.
4. **Pragmatic analysis**, the final stage, produces a complete meaning for the sentence via the application of contextual information. This information includes data that has to do with the time and location of the utterance, who was saying it, and to whom it was said.

Speech Recognition

Speech recognition by machine is a laudable aspiration. Wouldn't it be nice to be able to talk to our computers instead of having to type in commands or use a mouse? Humans use language quickly and effortlessly to communicate ideas to one another. To be able to communicate in a similar way with computers would usher in a new age of efficiency and productivity. Unfortunately, the task of getting a machine to understand speech is much more difficult than it may seem. Let's review some of the steps that speech recognition by machine would have to include and talk about the problems involved.

Any attempt at speech recognition starts with a **speech spectrogram**. A speech spectrogram is a visual representation of the speech signal; it is a graph that displays the component frequencies of a speech sound over time (see Figure 9.7). From this, a computer program then attempts to extract the phonemes from the segment of speech under analysis. If a phoneme is ambiguous, the segment of the speech signal that it occupies can be matched against similar utterances that have been recorded and analyzed to "fill it in." The phonemes are then assembled into their corresponding words. This is accomplished in part by a statistical analysis that factors in the probabilities that specific words will crop up in speech, that specific phonemes will crop up in specific words, and that specific words will be surrounded by other specific words.

Phoneme-to-word assignment is difficult for two main reasons. The first of these concerns word boundaries. In turns out that there are no pauses in-between words in spoken speech. This makes it hard to tell where one word starts and another ends. To compound the problem, there are often pauses within words. So pauses cannot serve as reliable indicators of word

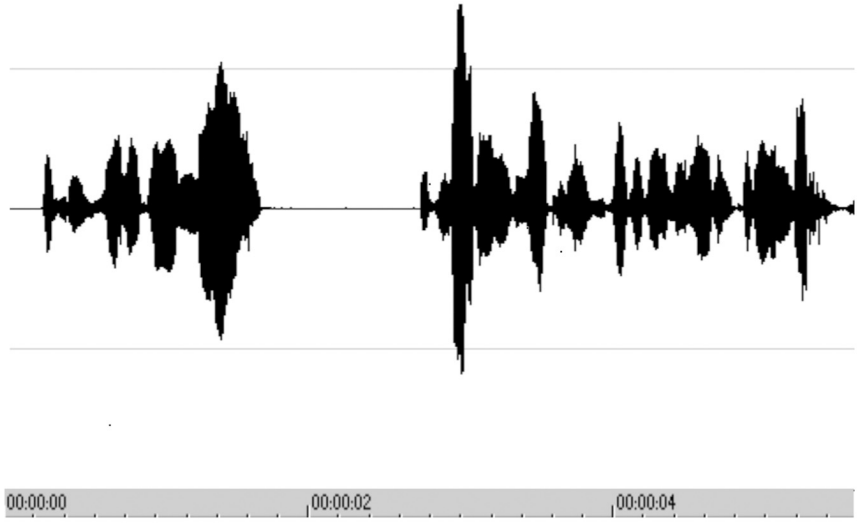


Figure 9.7 A speech spectrogram

boundaries. The second major issue is phoneme variability. If each phoneme were pronounced clearly and uniformly, speech recognition would be much easier. This is not the case. Speakers vary tremendously with respect to the pitches and durations of their phonemes. They are apt to pronounce a given phoneme variably, and in ways that depend on which phonemes precede it and which come after it. This is known as **coarticulation**. Additional complicating factors are the presence of background noise and the fact that in English a single sound, for example, that represented by "bear" and "bare," can belong to more than one word.

People resolve these difficulties by taking into account the overall meaning of a sentence. In one classic study, Warren and Warren (1970) presented participants with recordings of sentences in which a cough sound was substituted for a phoneme. One of the sentences was "It was found that the *eel was on the axle. (The asterisk indicates where the cough sound was inserted.) In other versions of the sentence, the word "axle" was changed to "shoe," "orange," and "table." Asked to interpret the four sentences, the subjects heard the ambiguous word as "wheel," "heel," "peel," and "meal," respectively. This clearly demonstrates that the meanings of words in a sentence that have already been understood provide a framework for understanding the words that have yet to be understood. It also shows that in human speech perception,

recognition is top-down as well as bottom-up, as the meaning of the entire sentence is pulled together simultaneously with the meanings of individual words.

Humans also have the benefit of visual cues when they are taking in speech. We can look at a speaker's lips as he or she is speaking. The positioning of the lips can help us to interpret difficult phonemes or morphemes. Because some deaf individuals can understand speech by way of lip reading, there is obviously more than enough information in the visual aspect of speech to enable comprehension. Most computer speech recognition systems in use today must rely on auditory information as their only input and so do not have visual cues. The In Depth section describes a cognitive model of how humans might recognize words presented to them either visually or verbally, get at their meanings, and produce responses.

Syntactic Analysis

Once the individual words and their order have been determined we can analyze the speech stream at the sentence level. This analysis entails the use of grammar. We have already discussed the various grammars that govern sentence structure. Syntactical analysis programs perform the equivalent of using a phrase-structure grammar to evaluate a sentence and to break it down into its hierarchical constituents. An understanding of this structure is necessary if we are to get at the sentence's meaning.

Semantic Analysis

The string of phonemes that make up a word are sometimes enough to reveal the word's meaning. This is achieved by a comparison of the phonemic string to an internal database of sounds. If a match is obtained, the word's meaning is derived. But many times, there isn't a perfect match and much ambiguity remains as to a word's meaning. In this case, the syntactical structure of the sentence can be useful.

In **compositional semantics**, the entire meaning of a sentence is derived from the meanings of its parts. A syntactical analysis identifies the type of word for each word in the sentence. This gives us information about how those words are related. For example, if a given word is identified as a noun, we know that it can be an agent of or instigator of an action. If a word is identified as a verb, we know that it represents an action. The structure of the sentence can then tell us whether that noun was the agent of that action. If a verb phrase is linked at the next higher level in the phrase structure to a noun phrase that immediately precedes it, there is a good chance that the action represented by the verb is brought into being by that noun. This structure has thus told us that the noun is an

agent—additional information that has to do with the word's meaning. Similarly, it is clear that the adjective inside a noun phrase is a descriptor that applies to that noun and not any other. The structure has again helped us to decipher meaning. We know that this word describes the noun in the phrase and not any other.

Consider the following:

"'Twas brillig, and the slithey toves did gyre and gimble in the wabe."

What makes this sentence interesting is that it uses imaginary words, yet follows proper grammatical rules. Even though most of these words are meaningless, we can still glean some understanding of what is happening by drawing on our knowledge of grammatical construction. You have probably guessed that the "toves" are the noun and "slithey" is the adjective that describes them. You may also have intuited that "gyre" and "gimble" are verbs that describe the actions performed by the "toves" and that they are doing it in the "wabe." Now consider this:

"Colorless green ideas sleep furiously."

Here we have an example of another sentence that follows grammatical rules. It differs from the one above in that it is composed of meaningful words. This time, however, the words are used in a nonsensical way. Ideas cannot be green, and if they could be, they could not also be colorless. Likewise, ideas can't sleep, and if they could, they couldn't do it furiously. This sentence is even more confusing than the one above because despite the fact that the words fit the "slots" that make up a proper phrase structure, their meanings conflict with one another.

So we see that grammar does more than tell us what a correct or incorrect sentence is. Grammar specifies the abstract relationships between words in a sentence. These relationships are important clues in our deciphering of individual word and overall sentence meaning.

Pragmatic Analysis

Human language is a social construct. Individuals use it to communicate with one another. We communicate for all sorts of reasons. Sometimes the purpose of a linguistic utterance is simple conveyance of information, as when we say "This pillow is soft." But linguistic utterances can serve many other purposes—ones that require an action on the part of the listener, for example. It is sometimes not apparent what action should be undertaken by the listener, as many sentences invite action on the listener's part without directly issuing a

command. **Pragmatics** are the social rules that underlie language use, as well as the strategies used by speakers to make themselves clear. Pragmatics helps us to understand what actions we should take in response to spoken sentences.

Searle (1979) outlines five different types of spoken statements. Each type demands a different response from the listener.

1. **Assertives** are spoken statements in which the speaker asserts his or her belief. An example would be: "It is hot in here." The statement suggests that we should open the window or turn on a fan or an air conditioner.
2. **Directives** are instructions dispatched from the speaker to the listener. They are direct commands and don't require an inference on the part of the listener with respect to what the desired action is. "Turn down the radio" is such a command.
3. **Commissives** commit the speaker to a later action, as when a child says, "I will take out the garbage later." We would then need to verify that the garbage had indeed been taken out or had not, and we would possibly impose a reward or punishment, depending on the outcome.
4. **Expressives** describe the psychological states of the speaker. "I apologize for yelling at you" indicates sorrow or regret, and implies that the speaker probably continues to trust the person being spoken to.
5. **Declaratives** are spoken statements in which the utterance itself is the action. "You are fired" means we have to look for a new job.

In each of these statements, we see that a speaker has used language to get the listener to perform an action. This is the case even in instances in which the sentence has not been phrased specifically as a command. Understanding the meaning of a statement is not enough here: one must infer what action the statement has asked for (directly or indirectly). Social context plays an important role in this process of establishing intent. "Do you know the time?" asked by someone waiting at the bus stop means that the speaker genuinely wants to know the time to find out if the bus is late. This same question asked of someone who has arrived late to an important meeting has a different intent. It is a criticism of that someone's lateness and not a request for the time.

Evaluation of Natural Language Processing

Computers are good at dealing with formal languages wherein everything is specified. They are not so good at dealing with natural languages that can be inherently ambiguous. Although recent years have seen advances in the

creation of machines that can understand speech, we still have a long way to go before we will be able to talk to our computer the way that Captain Kirk talks to the spaceship's computer in the TV show *Star Trek*. Speech recognition systems still have problems in dealing with noise, differences in pronunciation, and word ambiguity. Also, semantic analysis cannot rely on grammatical structure alone in the decipherment of meaning. This analysis must take into account real-world knowledge and social context. These bodies of knowledge are extensive and enormous. They include knowledge of objects, their uses, and the actions and lives of people, living and deceased. The ability to understand speech in a general way thus requires a very large database. Currently, speech recognition systems are only capable of understanding much more limited domains.

Overall Evaluation of the Linguistic Approach

Hopefully, this chapter has given you a sense of the importance and complexity of language. Linguistics resembles cognitive science as a whole in that it brings multiple perspectives to bear on a single topic. Instead of mind, its topic is language. Linguistics uses a varied set of theories and methods in its attempt to answer questions about language. The common thread that runs through its investigations is the subject matter, and not any one technique or orientation.

We have made great strides in expanding our understanding of language in the past few decades. Grammatic analysis has been used to expand our comprehension of language structure. We have elucidated the linguistic abilities of animals and we know the developmental stages that all children pass through as they acquire language. But there is yet much to learn. The relationship between language and thought is still murky, and we await further research that will shed light on the ways in which language influences thinking. The Wernicke-Geschwind model will in all likelihood be revised in favor of a new, comprehensive model that will be more detailed and will be able to specify the neural basis of processes such as retrieval from the mental lexicon. There will also be advances in computer-based language comprehension.

In Depth: The Logogen Model of Word Recognition

How is it that we recognize words? Every day we perform this task rapidly and effortlessly, while reading and while listening to the sound of a spoken voice. Yet this process is complicated and involves a number of processing steps.

Several models have been advanced to explain word recognition. One of these is a network model called TRACE (McClelland & Elman, 1986). Because it is in some respects similar to the NETtalk model, described previously, we will not discuss it here. Instead we will sketch out a long-standing cognitive information-processing model of word recognition, known as the logogen model (Morton, 1969, 1979). We will then contrast this model with more current neurological studies of the lexicon.

A **logogen** is a word representation. In the original formulation of this model, there is a logogen for every word. That means there is a logogen that corresponds to the word "elephant" and another that corresponds to the word "wet," and so on. These logogens are stored in a **lexicon**, a mental dictionary in which specific word features are represented. Morton (1979) specifies two lexicons—one for auditory word representations and another for orthographic or visual word representations (see Figure 9.8). The logogens that make up these lexicons are activated by sensory inputs from the corresponding sensory modalities. Sounds taken in while one is listening to speech first undergo acoustic analysis. The result of this analysis then triggers the activation of logogens in the auditory input lexicon. When one reads, the sight of words, after undergoing orthographic analysis, triggers the activation of logogens in the orthographic input lexicon.

Sometimes one's perceiving a word is insufficient to call forth its meaning. In these instances, contextual information, such as the meanings of the surrounding words and general world knowledge, is called into action. We referred to this earlier as representing the difference between bottom-up and top-down processing. In the logogen model, these two-way influences on word recognition are mediated not only by relations between words in the sensory inputs, but also by world knowledge that resides in the cognitive system. "Cognitive system" refers to other systems such as long-term memory. In Figure 9.8, arrows that lie between the cognitive system and the input lexicons point in both directions, indicating that both bottom-up (perceptual) and top-down (conceptual) information can drive the activation of word representations.

We have described the comprehension aspects of the logogen model and have shown how words are recognized while one is listening to speech or reading. The model can also explain production—how words are used to generate language during the operation of talking. This is achieved via a logogen output system. Activation of output logogens produces a phonological code. This code is fed to an articulatory apparatus, which enables the production of speech. One's thinking of something to say would activate the output logogens. This is represented in the model by the arrow that originates from the cognitive system and points to the logogen output system. Morton also speculates that output logogens can

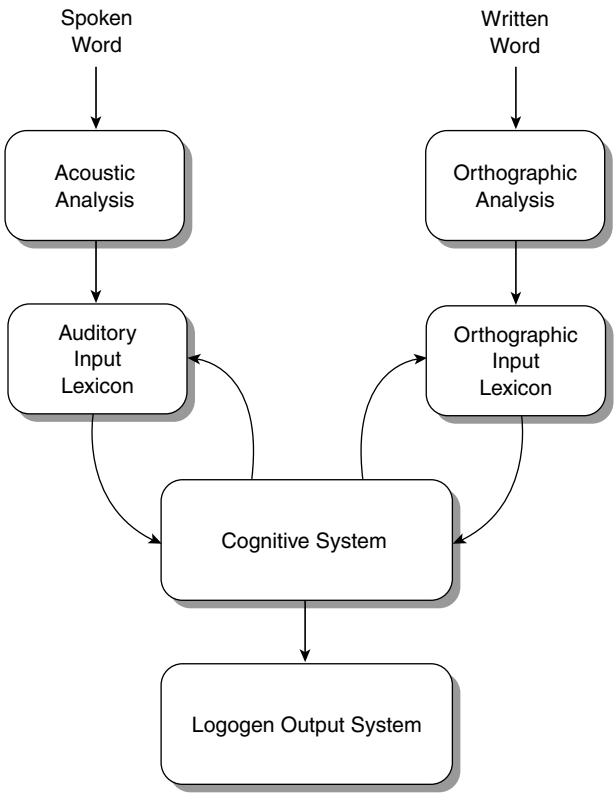


Figure 9.8 The logogen model of word recognition

be activated directly by their input analogs in the lexicons. This could occur, for example, when we are listening to what someone has just said and we make an automatic response. The arrows originating from the lexicons and pointing to the output system are used to represent this type of occurrence.

It is important to note that the activation of logogens does not necessarily call forth their meanings. In other words, the logogen for a particular word might be active in an individual's consciousness, and he or she might still not know its meaning. This is the case when we recognize an unfamiliar word, but don't know what it stands for. In this model information about word meaning is contained in the cognitive system, not in the logogens. This meaning (along with sensory inputs) is what influences word recognition.

To illustrate how the model works, consider the following sentence: "The carp swam in the pond." The logogen for "carp" that resides in the

orthographic input lexicon would be activated first by one's encounter with the printed word. Let's assume that you have seen the word carp before but don't know what it means. Upon continued reading, the meanings of other words in the sentence would help to disambiguate "carp." The word "swam" indicates that "carp" is probably a fish, since fish swim. The word "pond" would also suggest that it is a fish, since fish are often found in ponds. The utilization of one's knowledge of swimming and ponds represents the functioning of top-down activation within the cognitive system that helps to stimulate the carp logogen.

Modern Conceptions of the Lexicon

The logogen model is a nice way of conceptualizing some of the important issues in word recognition. It accounts for speech and reading comprehension via its postulation of two distinct pathways for the processing of acoustic and orthographic inputs. It takes into consideration the top-down and bottom-up effects that derive from a two-way interaction between the cognitive system and sensory inputs. It also explains how word representations are used in the output system that enables language production. However, there have been many studies of the lexicon and word recognition since the formulation and subsequent revision of this model. We will look at some of this work as a way of helping us to evaluate the older logogen ideas.

In cognitive science, the lexicon is considered to be a collection of the morphemes, both words and word-parts, that make up a language. Each morphemic entry is believed to contain three bits of information (Grimshaw, 1999). The first is semantic. It is the meaning of the morpheme and can be thought of as its definition; a morpheme within a lexicon has its definition in much the same way that a dictionary has a definition for every word. The second is the expression of the morpheme's syntactical function, for instance, how it is used in a sentence. Syntax here includes the kinds of grammatical relations the morpheme has with other morphemes. Finally, the morphemic entry is thought to possess morphological or phonological information that specifies its pronunciation. In the next few paragraphs, we will consider the neurological evidence that shows that the lexicon contains these and other types of information.

Patients who present with the aphasic deficit known as **anomia** are unable to remember the names of things, people, or places, but have some ability to define them (Kay & Ellis, 1987). This shows that the brain damage that these patients have sustained has disturbed the "labels" or forms of words but not their meanings. Conversely, other brain-damaged patients cannot define words, but their remaining lexical knowledge is intact, demonstrating that only the

semantic aspects of entries are affected. Both of these groups show evidence of widespread damage to their left temporal, parietal, and frontal lobes.

Yet another set of patients have difficulty with the spelling of words (the orthographic forms) or with their phonological characteristics. These people have no difficulty with word meaning. This suggests that these two attributes, spelling and phonological characteristics, are coded for separately and independently of semantic properties. These data provide some support for the logogen model, which postulates that the phonological, orthographic, and semantic components of words are parts of separate systems.

The deficits seen in these types of aphasias can be even more specific. There are patients who have difficulty processing words that represent items belonging to distinct categories, such as animals, plants, and constructed artifacts (Caramazza & Shelton, 1998; Gainotti & Silveri, 1996; Martin, Wiggs, Ungerleider & Haxby, 1996). For instance, a patient might be able to say "giraffe" upon his or her seeing a picture of that animal. Additionally, they would have no trouble identifying other animals represented in pictures. But this patient might not be able to identify pictures of tools and could have difficult identifying a drawing of a hammer or screwdriver. Additional category-specific deficits have also been found for nouns and verbs (Damasio & Tranel, 1993). This demonstrates that words that represent items or actions that belong to distinct categories are also coded for separately.

In conclusion, more recent investigations of the lexicon show us that it is distributed throughout the left hemisphere. Distinct areas of this hemisphere code for distinct aspects of lexical knowledge. These include the form of a word, its meaning, spelling, and sound. Furthermore, representations of words that belong to distinct word categories appear to be localized to distinct neural regions. These categories correspond broadly to living and non-living things and to noun-verb distinctions. Additional research will probably yield even more refined representational distinctions.

Minds On Exercise: Conversational Pragmatics

This exercise requires three people. Two will sit facing each other and engage in a five-minute conversation. A third will monitor the conversation and record the number of sentences that fall into the five categories of speech outlined by Searle (1979). Alternately, the observer can record the conversation and make a more careful transcription of the conversation at a later time. What types of speech were used the most? The least? Why? Were there any sentences that one partner in a conversational pair found difficult to comprehend? How did the speaker clarify his or her intent?

Food for Thought: Discussion Questions

1. Imagine that you have been trying to teach a chimp named Bongo to communicate with you. What kind of sentences could Bongo produce? What kind of sentences would he be unable to produce? Why?
2. What are the three stages that a child might go through in coming up with the past tense form of the verb "to run"? Give an explanation for each stage.
3. Draw a tree diagram for the phrase structure of the following sentence: "The police chased the men who robbed the bank." Do the same for: "The men who robbed the bank were chased by the police." Are the phrase structures different? Do the sentences have the same meaning?
4. Sketch out the components of the Wernicke-Geschwind model and the pathways that mediate information flow between them. Now go through and knock out one region at a time, simulating the damage that might be sustained by that area as a result of stroke. Describe the resulting signs and symptoms. For example, what kind of language deficits would a patient with damage to the angular gyrus have? In each case, describe why the patient would exhibit those deficits.
5. Make a telephone call to an automated voice recognition service. These are commonly used by some customer service departments of telephone companies and by services that sell movie tickets. In what ways is the system rigid? Why is it so difficult to get a computer to process natural language?

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Allen, J. F. (1995). *Natural language understanding*. Redwood City, CA: Benjamin/Cummings.

Carroll, J. B. (Ed.). (2000). *Language thought and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

Lane, H. (1976). *The wild boy of Aveyron*. Cambridge, MA: Harvard University Press.

Rymer, R. (1993). *Genie: A scientific tragedy*. New York: Harper Collins.

Savage-Rumbaugh, E. S., & Lewin, R. (1994). *Kanzi: The ape at the brink of the human mind*. New York: Wiley.

Terrace, H. S. (1979). *Nim. A chimpanzee who learned sign language*. New York: Columbia University Press.

10

Artificial Intelligence I: Definitional Perspective

"Artificial stupidity may be defined as the attempt by computer scientists to create computer programs of a type normally associated with human thought."

Wallace Marshal

Introduction

I propose to consider the question "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think."

With this provocative question and challenge, Alan Turing began his seminal paper *Computing Machinery and Intelligence* (1950). Turing's proposal evolved into what he called an ultimate test of intelligence. If a machine were able to pass the test—discussed in detail in this chapter—in its most refined form, it would enable one to make the argument that the "intellectual" responses of a machine were indistinguishable from those of a human being. Turing's contributions have produced an avalanche of passionate debate and research in the 50 years since the signal moment at which he so eloquently

articulated the central questions of Artificial Intelligence (AI). In this chapter we will bring together the elements of technology and behavioral and cognitive models that provide the basis for the possibility of creating a machine-based intelligent agent—a goal that has intrigued humanity for several millennia. In short, we seek to develop a definition of AI.

To accomplish this goal, we will briefly review the philosophical and technological streams whose convergence now comprises the basis for AI. The concepts of strong AI, applied AI, and cognitive simulation—and the distinctions between them—will be defined and demarcated. The use of the characteristics of intelligence as a cornerstone of AI will be explored, as will human learning paradigms. The two fundamental approaches to AI—top-down and bottom-up—will be defined. The digital computer as the foundation for the modern intelligent agent will be introduced, including an explanation of its "linguistic" or programming strictures.

Turing will be revisited in considerably more detail, as will the arguments of his detractors in this ongoing debate regarding machine intelligence and the "battle lines" of the current disputes on the subject of the future of AI. As in other chapters in this text, an In Depth discussion will be provided—in this instance we consider the contributions of Ned Block and behaviorism.

Historical and Philosophical Roots

Witness the fascination that we have for store windows full of mechanical displays, particularly around the Christmas holiday season. Often, very elaborate scenarios and stories are played out by innovative, humanlike **automata** displaying motions that parallel our own behavior. As applied here, the term "automata" refers to mechanisms that can move or act of themselves. One of the most popular robotic "toys" of 2005 was "Robosapien"—a computer-based automaton that, according to its creators, achieves fast, dynamic, two-speed walking and turning, performs 67 pre-programmed functions (e.g., kicks, dances, and exhibits several scatological capabilities), is fully programmable by remote control, and is capable of fluid "international caveman speech." Although we do not consider such toys to be human, they are demonstrations of our long-held desire to create a machine "in our own likeness"—an automaton that would exhibit such lifelike qualities that we would no longer be able to discern any differences between machine and human behaviors. The advent of the electronic digital computer has altered the focus of our search from the mechanical to the intellectual capacities of humans.

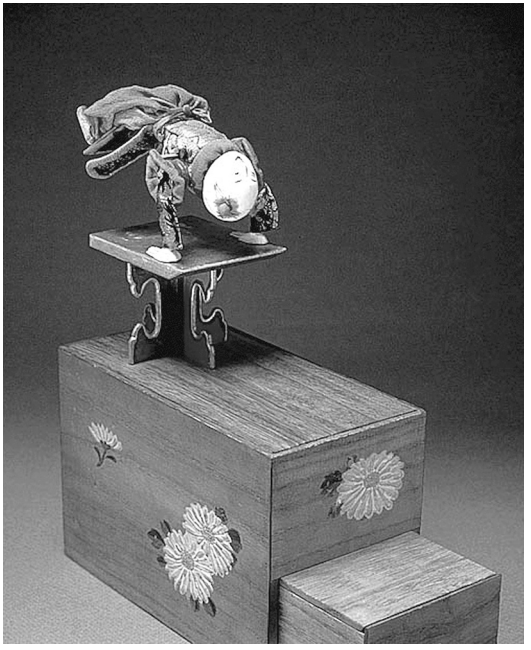


Figure 10.1 Mechanical somersault doll, eighteenth century, Edo Period

Source: Reprinted with permission of the Kyoto Museum.

The Quest for "Mechanical Life"

AI would not be possible without the advancements accruing from mechanical technology. The history of mechanical dolls can be traced back at least 2,000 years, to a time when the Greeks used wind and water to power moving statues and introduced manmade singing birds. During medieval times, mechanical birds and moving dolls appeared in India and Arabia. People in these places viewed them as magical and fearsome, while religious groups condemned them as devilish and sought to banish them from daily life. Some examples from history include:

- Louis XII of France receives a mechanical lion from Leonardo Da Vinci (1509). It can move the length of a long hall, stop in front of the king, and place a fleur-de-lys at his feet.
- King Gustavus Adolphus of Sweden receives a cabinet in which two well-dressed dolls dance together (1632).

- Louis XIV receives a miniature horse-drawn carriage complete with attendant servants. It travels over a table, stops in front of the king; a doll emerges and presents the king with a petition. The doll returns to the carriage, which then drives off.
- Pierre Joquet-Droz is invited to the court of Louis XVI and Marie Antoinette to display "The Writer," which can write sentences. Another automaton can draw pictures, and yet another can play the clavichord.

The eighteenth century was perhaps the golden age of the "philosophical toy." And, one of the most colorful technicians of this age was Jacques de Vaucanson. His magnificent creations were admired all over Europe. The French philosopher Voltaire referred to him as a "new Prometheus." When de Vaucanson was very young, his mother would take him to church, where he was left alone while she went to confession. To pass the time, the young boy noted every detail of a clock in an adjoining room and he produced a perfect copy of the mechanism at home. He eventually went on to study with a Jesuit order (l'Ordre des Minimes) but ultimately found this life to be too confining and he fashioned a profitable livelihood from his automata. A brief list of his mechanical constructions would include the following:

- The Flute Player: a mechanism that could play the flute. The automaton was fashioned after a famous statue by Antoine Coysevox; it was life-size, a tune was generated by fingers and breath was produced by varying amounts of air blown into the flute in a manner that depended on the shaping of the lips. To produce the sounds, nine bellows were attached to the chest. The bellows were attached to different weights so as to give out varying degrees of air. The pipes were conjoined in a kind of trachea and from there the air passage widened to form a throat and mouth cavity. The machine could play 12 different melodies. It mimicked the means by which a man would make the sounds and there was a mechanism to correspond to each relevant muscle.
- Pipe-and-drum figure: it played a pipe at a speed faster than any living person could achieve.
- The incontinent duck: his most famous device. This remarkable duck ate food from an exhibitor, swallowed it, digested it, and, yes, produced excreta. Voltaire thought that this contribution echoed the glory of France. It was similar in size to a duck, drank water with its beak, quacked, could rise up and settle back on its legs, and, most amazingly, it swallowed food with a realistic gulping action of its flexible neck.

Because Louis XV was an admirer of the duck, he appointed de Vaucanson an inspector of silk manufacture. De Vaucanson set to work automating the looms of the region, and as a result there was great upheaval in the silk mills of Lyon. In retaliation against the scorn of the millworkers, he built a loom that could be operated by a donkey—to prove that "a horse, an ox or an ass can make cloth more beautiful and much more perfect than the most able silk workers." The anecdote provides an example of an instance in which humans were threatened by the possibility of their replacement by machines—an issue of increasing importance for contemporary societies in which machines are increasingly capable of completing tasks that have been considered the exclusive domain of humans.

In England, automata began to appear for general public consumption during the Victorian Era, roughly 1837–1901, corresponding to the reign of Queen Victoria. Mechanical monkeys, drinking tea and smoking, were displayed in shop windows, to attract customers. By 1880, small steam engines and battery-operated boats as well as trains and fire engines were much in evidence.

Philosophical Origins—Man as a Machine

The philosophical distinction between mind and body was first introduced in Chapter 2 (The Philosophical Approach), where it was explored in detail. While the debate can be traced to the Greeks, it is in the seminal work of René Descartes (1596–1650) that we find the first systematic account of the mind/body relationship. Descartes was known to have attempted to build automata in his life. In connection with this, the philosophical questions introduced by Descartes are: "What is the difference between a person and a machine? Where is the line between the animate and the inanimate, between life and death? Is there a difference between reason and 'randomness'?"

Are such questions simply matters of philosophy? Until the nineteenth century, "philosophy" included all branches of science. But science has moved us from an age of alchemy to one in which alchemy and other magic arts have been replaced by mechanics, astronomy, and so on, and in which scientific information is abundant and widely available. (It is reported that Descartes "entertained" the sick with mathematics.) The modern equivalent of these developments can be found in academic courses that have titles such as "Physics for Poets"—courses that enjoy considerable popularity on college campuses. What came about as the central idea of Descartes' age was simply that man is a machine and should be understood as such. Descartes' "Treatise on Man" includes a comparison between a human being and a hypothetical "statue or machine" that operates like a clock or a hydraulic fountain. The

difference, as proposed by Descartes, is that a human being possesses a "rational soul," whereas an animal is not capable of reasoning. His famous quotation, "I think therefore I am," epitomizes this difference.

Evaluating Descartes' Approach

Descartes evoked an opposing principle in the mind/body debate. The atheist philosophers of the eighteenth century took great issue with the idea that a "soul" separates us from the machinelike model that describes other animals. Consider Jaquet-Droz's automaton (currently in NeuchActel, Switzerland), described as being able to write "I think therefore I am." Yet, it is also capable of coming up with a more ironic tribute to Descartes: "I do not think . . . do I therefore not exist?" The writer (automaton) is able to declare that it cannot think. If it can communicate that it cannot think, is it possible that it really can think after all? Is the machine lying—a most human activity?

In his treatise, Descartes carefully avoids the conflict; the machine is not a man, only a "statue or machine . . . which God forms with the explicit intention of making it as much as possible like us." Descartes' machine includes both a body and a soul—he intended to describe each separately—but the portion of the treatise that deals with the soul has been lost, and only a mechanistic interpretation survives:

> "I desire . . . that you should consider that these functions follow in this machine simply from the dispositions of the organs as wholly natural as the movements in a clock or other automaton follow from the disposition of its counterweights and wheels."

The overarching quest to imitate human beings stretches to the arts. Standing out among numerous works of art and art forms, the opera "The Tales of Hoffman" by Jacques Offenbach tells the story of the poet Hoffman and his ill-fated three loves. His first love, Olympia, is a mechanical doll created by Coppelius (an incarnation of the devil). Hoffman falls in love with the doll, which is eventually smashed by its angry creator.

Mechanical Computation

The history of AI, which includes the philosophical considerations proposed by Descartes, would not be complete without consideration of mechanical computation. A part of human intelligence rests on our ability to perform

Table 10.1 Events in the Development of Mechanical Computation

Year	*Event*
30,000 B.C.E.	Paleolithic people in central Europe record numbers by notching tallies.
2600 B.C.E.	The Chinese introduce the abacus. (Used in China as recently as 1982.)
1500	Leonardo da Vinci invents the mechanical calculator.
1642	Blaise Pascal invents the first usable, hand-turned calculator for addition and subtraction (which he names *Pascaline*).
1670	Gottfried Leibniz enhances Pascaline with multiplication, division, and square root functions.
1822	Charles Babbage invents the Difference Engine (mechanical calculator), capable of adding and subtracting.
1936	Konrad Zuse creates a programmable digital computing machine (with vacuum tubes and a binary number–based architecture).

calculations (as distinct from our ability to abstract information and represent it in numerical form). Interest in machine computation has, in part, paralleled the development of mechanical dolls. Table 10.1 traces a few of the developments in the history of mechanical computation.

As noted in the table, primitive but seminal mechanical measurement is traceable back to more than 30,000 years ago. Improvements in counting—a significant part of mechanical computation—were embodied in the Chinese abacus (ca. 2600 B.C.E.). Calculators that incorporate a measure of automation can be traced to Leonardo da Vinci (with later contributions made by Pascal and Leibniz) and particularly to Charles Babbage. The modern age of automation received an enormous boost from the capabilities of the electronic vacuum tube (and subsequently from the transistor and the integrated circuit).

Mechanical calculators are considered to be automata capable of "imitating" human intelligence on what one could call a "primitive" level. Two examples of such calculators are shown below. The abacus is shown in Figure 10.2, and a fragment of Charles Babbage's Difference Engine is shown in Figure 10.3.

The abacus was developed some 4,000 years ago. In some remote parts of China and Japan such calculators are still in use. One can also find these "cash registers" in some grocery stores within the Chinese neighborhoods of large cities in the United States. Well-trained abacus users will often have better results than those using a mechanical or electrical calculator, when both are adding the same column of numbers.

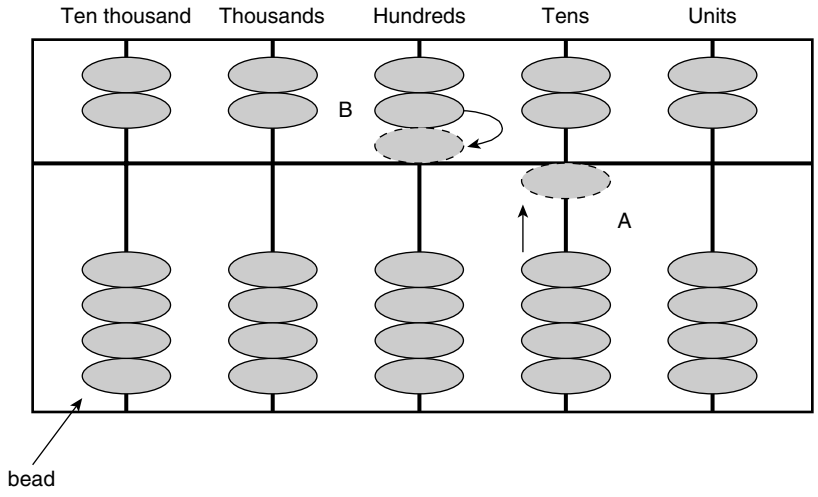


Figure 10.2 A sketch of one form of an abacus indicating how decimal numbers are represented. Each bead below the bar represents 1; each bead above the bar represents 5. In addition, beads are weighted according to their decimal position. Thus bead A, when moved to the bar, has a value of 10, while bead B has a value of 500. The number represented by the dotted-line beads is therefore 510

Some argue that Charles Babbage actually conceived the "digital" computer that dominates our daily activities of living. Babbage (1792–1871) was an eminent scientist of his day. He was the Lucasian Professor of Mathematics at the University of Cambridge. This Chair has also been held by such notables as Sir Isaac Newton and, most recently, Stephen Hawking. Babbage first developed a mechanical calculator, the **Difference Engine**. He subsequently designed, but never built, an "**Analytic Engine**," whose architecture (or organization) greatly resembles the modern **personal computer** (PC). The Difference Engine (1821) was the world's first programmable, automatic, digital calculating machine. The only human intervention that was required was the setting of initial parameters (constants), followed by the turning of its handle. It was designed to solve mathematical equations known as polynomial equations. Such equations were used in navigation systems (whose calculations, when made by hand, were prone to error). The Difference Engine used a numerical technique known as the Method of Differences. (A discussion of this technique is beyond the scope of this book.) The design of the Analytical Engine, according to many, anticipated virtually every aspect of present-day computers. Babbage envisioned

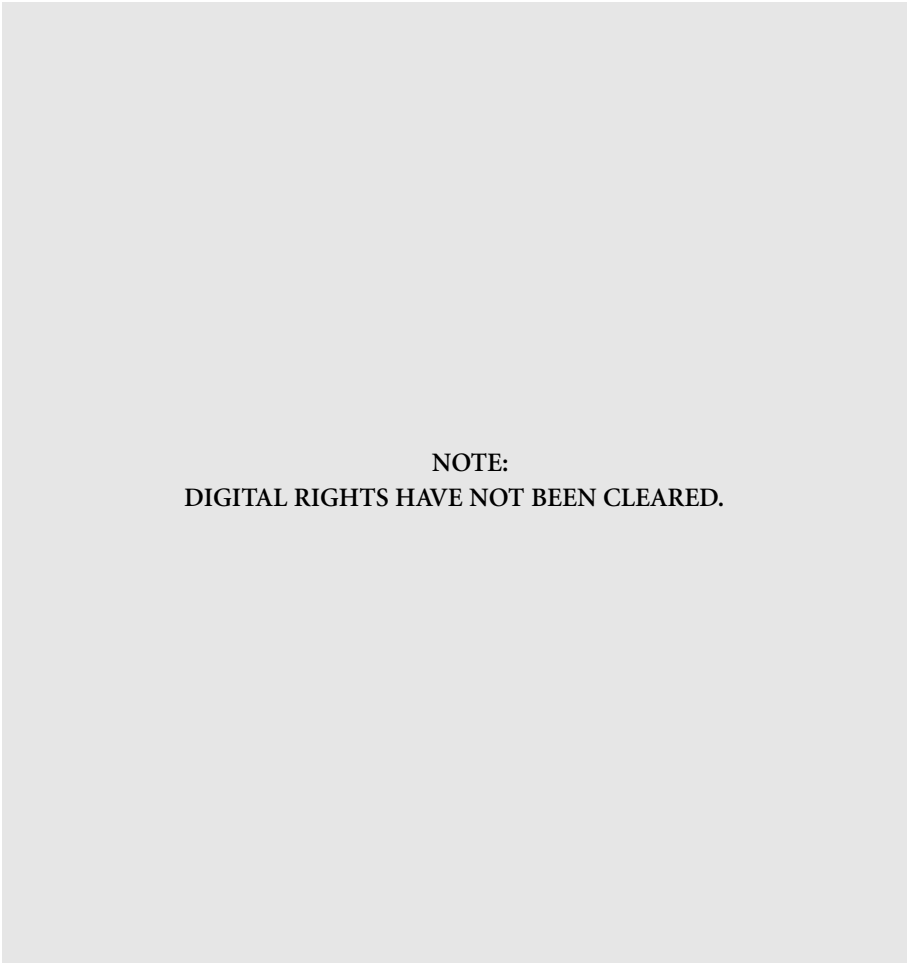


Figure 10.3 Photo of Charles Babbage's Difference Engine. This is only a fragment, as the complete calculator was never built.

Source: Reproduced with permission of the Science & Society Picture Library, The Science Museum, London, England

a massive, steam-powered, general-purpose, mechanical computer. Regrettably, he failed to persuade prospective backers to provide the funds that would have been needed for its actual construction. Its relevance to the modern computer is reflected by the fact that one can participate in its actual operation by using an available **emulator**, written in the Java programming language; it works exactly as designed by Babbage and you can ask it to solve problems. (An emulator

is a program that allows software intended for one computer platform or one type of hardware to be run on another platform or type of hardware.) With the appropriate PC software, one can run the Analytic Engine emulator, found at www.fourmilab.ch/babbage/applet.html).

Defining Artificial Intelligence (AI)

In setting out to define artificial intelligence, we recognize three levels of intelligence. An ant has intelligence of a very primitive level, higher-order animals (and simple robots) can be considered to have a qualitatively higher level of intelligence. At the top of the intelligence pyramid is the human agent. It is a defining objective of AI scientists to build an automaton that is capable of behaviors that would correspond to intelligence at this highest level, such that the automaton would be indistinguishable from a human being. Definitions of AI are numerous and have their different emphases that support their authors' special interests. (Refer to the discussion of AI in Chapter 1.) Consider one discussion that is sometimes cited in arguments that center on the varying definitions of intelligence. To a great extent, airplanes are controlled by the "automatic pilot," but would we conclude that an automaton is flying the plane? How does one, or can one, distinguish between a highly deterministic control system and a human intellect? Externally, the control systems perform the same sorts of actions that human beings would perform in similar circumstances. In fact, in some circumstances the automatic systems perform better than human operators. With respect to the limited environments in which they operate, do we conclude that these automatic systems have intellect equivalent to that of a human being? One could make analogies similar to this one that have to do with our activities of daily living.

Although we can identify a number of common themes within definitions of AI, each definition emphasizes an aspect of AI that its author considers important to his or her purposes. William Rapaport (of the Center for Cognitive Science at the State University of New York at Buffalo) has compiled a number of these definitions. A sampling of his definitions includes the following:

> "The goal of work in artificial intelligence is to build machines that perform tasks normally requiring human intelligence." (Nils J. Nilsson, 1971)

> " . . . Studying the structure of information and the structure of problem solving processes independently of applications and independently of its realization in animals or humans." (John McCarthy, 1974)

> "Artificial intelligence is concerned with the attempt to develop complex computer programs that will be capable of performing difficult cognitive tasks." (Basil Blackwell, 1990)

From these attempts at definition, it emerges that AI can be recognized as an important branch of engineering, of computer science—and as a distinct branch of cognitive science. *In short, there may not be any simple definition.* What are not provided in definitions of AI are the criteria that would enable us to know if an automaton exhibits "intelligence." In fact, one of the great debates of science is summed up in the question "what is intelligence?" A key element of this debate is to be able to determine whether or not an automaton represents "true intelligence," or whether it is simply a sophisticated machine capable of deterministic control of some logical process. (Many of the criticisms of the AI approaches and AI applications described in this text are central to this fundamental philosophical argument.) Some investigators consider automata to be nothing more than electronic equivalents of the mechanical dolls discussed previously. Whether or not an automaton is capable of things that would reflect true intelligence is a difficult question to resolve. Historical developments in the twenty-first century *may* resolve this question. To a great extent we rely on historical, "fashionable," or popular conclusions to come up with answers to this question. To move us closer to an answer, we emphasize the goals of AI science.

- The **engineering goal** of AI is to assemble systems that use a computer's facilities and inventory of knowledge in order to solve real-world problems.
- The (cognitive) **scientific goal** of AI is to codify knowledge, and **meta-knowledge**, in order to assemble systems that will shed light on and even explain various sorts of intelligence and consciousness. ("Meta-knowledge" refers to facts about knowledge itself, such as the rules we have for using such knowledge.)

Historically, if the **paradigm** on which the machine is crafted was related to a similar model within the human, it has been cast as an AI. (A paradigm is a set of assumptions and techniques that characterize an approach to a class of problems.) This principle has led to a classification scheme, according to which there are three general classes of AI: **Expert Systems, Fuzzy Logic, Neural Networks**. Expert Systems seek to emulate the abstract thinking and logic that humans use in order to solve problems; Fuzzy Logic (see Chapter 11) has origins similar to those of Expert Systems but seeks to account for the defects in human logic and the bad choices that human beings often make; Neural Nets (see Chapter 7) can be compared to a simple model of the neural system (the brain).

The potential uses for such automata are thought-provoking:

- In farming: computer-controlled robots that will be able to control pests, prune trees, or selectively harvest mixed crops.
- In manufacturing: AI-based robots that will perform ***d***angerous, ***d***ull, or ***d***irty ("3D") operations.
- In medicine: machine intelligence that will be able to diagnose disease, monitor patients, operate on patients (including telesurgery—performed by a surgeon at a distance), manage treatments, rehabilitate neurologically impaired individuals, make beds, provide care for the elderly and/or disabled.
- In activities of daily living (ADL): computers that will be able to advise on cooking and perform routine household chores.
- In education: computers that will be able to understand why students make mistakes (as opposed to simply reacting to those mistakes) and provide students with detailed information on topics taken from many disciplines, such as music, history, and the sciences.
- In research: machines that will be able to devise theories based on experimental observations.

These are all instances in which machine operation would have a striking (complementary) resemblance to human judgment, thinking, and logic. However, all of the preceding must be viewed as long-term objectives because we currently do not know how to design computers that will possess the comprehensive capabilities of perception and reasoning and the repertoire of actions of human beings. Some argue that some AI-based machines that exist presently do, in fact, possess the perception and reasoning capabilities of a human being, although they were not designed with these qualities in mind. For example, a neural net can "extract" the essential associative rules regarding input and response that humans are not able to discern.

Evaluating the Concept of AI

Can a mind be constructed? This is the critical question that best identifies the ultimate disposition of AI. AI research is roughly divided into three distinct categories: "strong" AI, applied AI, and cognitive simulation.

Strong AI: Advocates of this concept seek to build machines whose intellectual ability cannot be distinguished from that of a human being. Joseph Weizenbaum (of the MIT AI Laboratory) presents this view in its most compelling form. In his representation, the goal of AI is "nothing less than to build a machine on the model of a man, a robot that is to have its childhood,

to learn language as a child does, to gain its knowledge of the world by sensing the world through its own organs, and ultimately to contemplate the whole domain of human thought" (Weizenbaum, 1976). John Searle (of the University of California at Berkeley) introduced the term "Strong AI" in 1980; he is a strong critic of the Turing Test, briefly noted above and described in detail below (Searle, 1984). Many prominent theorists of the 1950s (e.g., Marvin Minsky and Seymour Papert) were excessively optimistic about achieving the ideal set forth in the strong AI model. Their optimism gradually receded, as they began to appreciate the extreme challenges to be surmounted (Minsky, 1982). Building a "true android" may be the most difficult technological undertaking ever attempted by science. An **android** is a **robot** or mechanical device that resembles a human being and performs automated tasks. (See Chapter 12 for a discussion of robots.) Some argue that we have not yet built a machine that possesses the intellectual capability of an ant, and that the prospect of an AI worthy of the designation of "human" remains an elusive dream. However, the dream is alive, and scientists such as Hans Moravec (Professor of Robotics at Carnegie Mellon University in Pittsburgh) and Rodney Brooks (Director of the Artificial Intelligence Laboratory at MIT) believe that a human being is nothing more than a fancy machine, and that as technology advances, it will be possible to build a machine with human features. Is there anything magical about the brain and biological flesh? As Moravec notes, "I'm confident we can build robots with behavior that is just as rich as human being behavior. You could quiz it as much as you like about its internal mental life and it would answer as any human being" (Brooks, 2002; Moravec, 1999). Others (e.g., David Chalmers, Professor of Philosophy and Director of the Center for Consciousness Studies at the University of Arizona) argue that consciousness is an inexpressible trait and it may be useless to try to pin it down. Chalmers, however, is open to the idea that machines "might eventually become conscious." Chalmers writes: "There may not be anything fundamentally different between a creature of flesh and blood and one of metal, plastics, and circuits" (Chalmers, 1997).

Applied AI: This concept of "intelligent" machines has enjoyed the most success. It is also referred to as "advanced information-processing." Its aim is to produce commercially sustainable "smart machines," some of which have been noted above. Chapter 11 focuses on key elements of applied AI.

Cognitive simulation and natural language communication: The focus within cognitive science (CS) is to understand how the mind works. CS was first introduced in Chapter 4 (The Cognitive Approach I), where it was explored at some length. Research in this area has included studies that have focused on how we recognize faces or other objects, or how we solve abstract problems. Within CS,

an algorithm in the form of a computer program is tested against the actions or thinking of a human being. Computer simulations have contributed to both psychology and neurophysiology. (See the discussions of connectionism as well as Neural Nets.) Consider the following results of a program generated by the Stanford psychiatrist Kenneth Colby that first appeared in 1966 (Colby, 1975). The program simulated responses by a human paranoiac. The program, called **Parry**, was capable of carrying on conversations that bore an eerie resemblance to real conversations, and psychiatrists who were asked to decide whether they were communicating with the machine (program) or a human paranoiac were often unable to. In what follows, a psychiatrist is generating questions that are presented to Parry by the "Interrogator," while Parry provides the responses:

Interrogator:	Why are you in the hospital?
Parry:	I shouldn't be here.
Interrogator:	Who brought you here?
Parry:	The police.
Interrogator:	What trouble did you have with the police?
Parry:	Cops don't do their job.
Interrogator:	What do they do instead?
Parry:	Cops arrest the wrong people.
Interrogator:	They arrest the right ones sometimes.
Parry:	The law should get those Italian crooks.
Interrogator:	Are you Italian?
Parry:	My background is British but I was born in this country.
Interrogator:	Do you know any Italian crooks?
Parry:	I try to avoid the underworld.

It is true that the Parry program depended on pre-constructed responses devised by the programmer and stored in the computer's memory. These were "finagled" by skilled computer programming in such a way that Parry was able to participate in dialog. The interaction noted above was able to "fool" a human expert.

No discussion of AI would be complete without considering the question: "Is Strong AI Possible?" We consider strong AI issues with the psychological mind-body discussion acting as a backdrop and describe the goals of strong AI.

An ability to use language, the ability to plan, and the ability to reason characterize human behaviors; these reflect underlying cognitive abilities. The

goal of strong AI is to duplicate these human intellectual abilities. One of the problems associated with this controversial goal has to do with the harvest from prior research efforts in this area—which has been disappointing, to say the least. Over the years, exaggerated claims of success have been reported in scientific journals as well as in the popular press. Although some success has been achieved in Robotics (see Chapter 12), after some five decades of research, attempts to "scale up" modest levels of "intelligent" machine performance (e.g., performance that has to do with obstacle avoidance, gaze control, and object manipulation—building structures made up of blocks) have failed to produce machines that live up to standards of human behavior to any significant degree. Connectionists cannot yet claim to have a working model of even the simplest of living things. For example, *Caenorhabditis elegans* is a worm with some 300 neurons and a known interconnectivity, yet connectionist models have failed to mimic this "simple" nervous system. Moreover, by the admission of the fabricators of these models, connectionist "neurons" are gross simplifications of the animal counterpart. While the actualization of strong AI may be terribly difficult, it remains a realistic objective for many researchers in AI.

Can a machine possibly be intelligent, think and understand? In the first instance, such a question may not be relevant; Noam Chomsky suggests that it is a question of decision, not fact. We must ultimately agree on what constitutes and defines intelligence, thinking, and understanding (Chomsky, 1968). As a simple illustration of the problem: we conclude that airplanes "fly" but that ships do not "swim," even if the ships are able to navigate in water just as animals (e.g., humans) do when they swim. While this may be an oversimplification, it is essential to determine the criteria and conditions that machines must satisfy in order to allow us to conclude that they can achieve the consciousness and understanding of the human model. As regards this point, we have not reached consensus. Some invoke the Turing Test (TT) as a "definition" of intelligence—a machine is intelligent if and only if the test fails to distinguish between human performance and machine performance. However, Turing himself noted that this test couldn't define intelligence. A computer could be considered to be intelligent in spite of the fact that it had failed such a test (i.e., simply because it failed to "imitate" a human being). Should we require that an extremely intelligent machine designed for a particular task—designed, say, to perform surgical operations with intelligence comparable to that of a skilled surgeon—meet the requirements of the TT?

Alternatively, if a computer should pass an "intelligence" test, would we impart intelligence to that machine? Claude Shannon and John McCarthy (1956) noted that one could "stack the deck": design a program with pre-programmed

responses to questions that the program would be apt to encounter while it was being put to the test. Like Parry, it would simply look up appropriate responses from its database. A number of people have championed this point of view (Ned Block, Stephen White, and Jack Copeland). A system with "no intelligence" at all might pass an arbitrary test similar to the TT.

Currently, AI has no universally accepted definition of intelligence to which it might anchor itself. Thus, there is no way of telling which research results do, in fact, demonstrate "intelligence." At the moment, such a definition is a "moving target" among professionals and non-professionals alike. Consider a program that can summarize newspaper articles, or a program that can beat a chess champion. Critics are quick to point out: "That's not intelligence"—although these are often the same critics who had previously maintained that a machine could "never do that" because it was not "smart enough."

One of the most eminent AI researchers, Marvin Minsky, points out that "intelligence" is our name for whichever problem-solving mental processes we do not yet understand. Making an analogy, he equates intelligence to the idea of "unexplored regions of Africa": such an unexplored region disappears when we discover it. Turing made the same point: "One might be tempted to define thinking as consisting of those mental processes that we don't understand." The problem of coming up with a clear set of criteria that would provide the basis for determining success in AI research remains. Can we even come up with criteria that would provide the basis for conclusive evidence of *human* intelligence?

Although several definitions of AI have been proposed, one broad definition describes AI as the science of making computers do things that require intelligence when done by humans. To understand the complexity of building such an automaton, some understanding of what is meant by "human intelligence" is necessary. In particular, there are several elements that seem to constitute the essential characteristics of intelligence: learning, reasoning, problem solving, perception, and language understanding. (Each of these is discussed in what follows.) AI machines whose design has focused on these components have garnered the most successes. These are among the same components that are considered in mainstream psychology to be essential to human intelligence.

Although relatively simple human behavior is considered to be intelligent, complex behaviors performed by insects are often designated as unintelligent. One species of wasp (*Shex ichneumoneus*) carries on an elaborate behavior when bringing food to its burrow; the behavior is designed to ensure that the activity will be completed successfully. The female wasp first deposits the food on the threshold of the burrow, goes inside to check for intruders, returns to the threshold, and, assuming that there have been no intruders, finally brings the food inside. If a researcher moves the food while the wasp is checking the

nest, she will repeat the entire procedure, starting from scratch. This behavioral response, which seems remarkable to us, can be repeated in excess of forty times. It appears that "intelligence" is conspicuously absent. A key element of intelligence is clearly missing, namely, the ability of the wasp to adapt its behavior to fit new circumstances.

Learning, in the context of a behaviorist or functional approach, can take several forms: **trial-and-error**; **rote learning**; **operant conditioning**. The simplest learning is **trial-and-error**. When a computer program has tried chess moves at random until it finds one that achieves checkmate, it will remember that move—such that the problem is immediately solved if and when the same circumstances present themselves a second time. This description of trial-and-error learning parallels a similar mechanism in humans. **Rote learning** involves direct association between a stimulus and a response; memorizing a list of items is an example of rote learning and is easy to implement on a computer. However, learning involves the ability to generalize in such a way that the intelligent agent (e.g., human or computer) will perform satisfactorily when a novel situation is encountered. For example, a computer can learn the past tenses of regular English verbs in a rote fashion. However, it would be impossible for the computer to generate the past tense of "jump" unless it had been presented with "jumped" at least once in the past. An "intelligent" program—one that can generalize—will employ the "add-ed" rule and produce the correct response even though it has not previously encountered the verb. AI machines have had reasonable successes with respect to this type of generalization.

Operant conditioning is a highly developed form of learning; additional discussion can be found in Chapter 3 (The Psychological Approach). Consider the case in which a human operator learns to slow a car when approaching a red light. Although it may first appear that this is a case of rote learning, further analysis reveals that it is something else: one should not immediately apply full pressure to the brake, as rote learning might imply. Rather, a sequence of learning steps is required such that one gradually develops the braking skill associated with an experienced driver. Automated operant conditioning systems are a relatively new development in AI. Automata have been developed to assist therapists in rehabilitating individuals afflicted with neurological disorders of movement, for example, stroke (Silverman, et al., 1991; Silverman, 2001). A generalized machine-based algorithm that can implement operant conditioning (i.e., positive and negative reinforcements) for learning behavioral skills is depicted in Figure 10.4. In this model, the mind and its deep complexities or its operational details are not part of knowledge/skill acquisition. Rather, the model reflects the views of behavioral psychologists. In this sense it has an "engineering" flavor, since engineers are often interested in modeling relationships between stimuli (inputs) and responses (outputs).

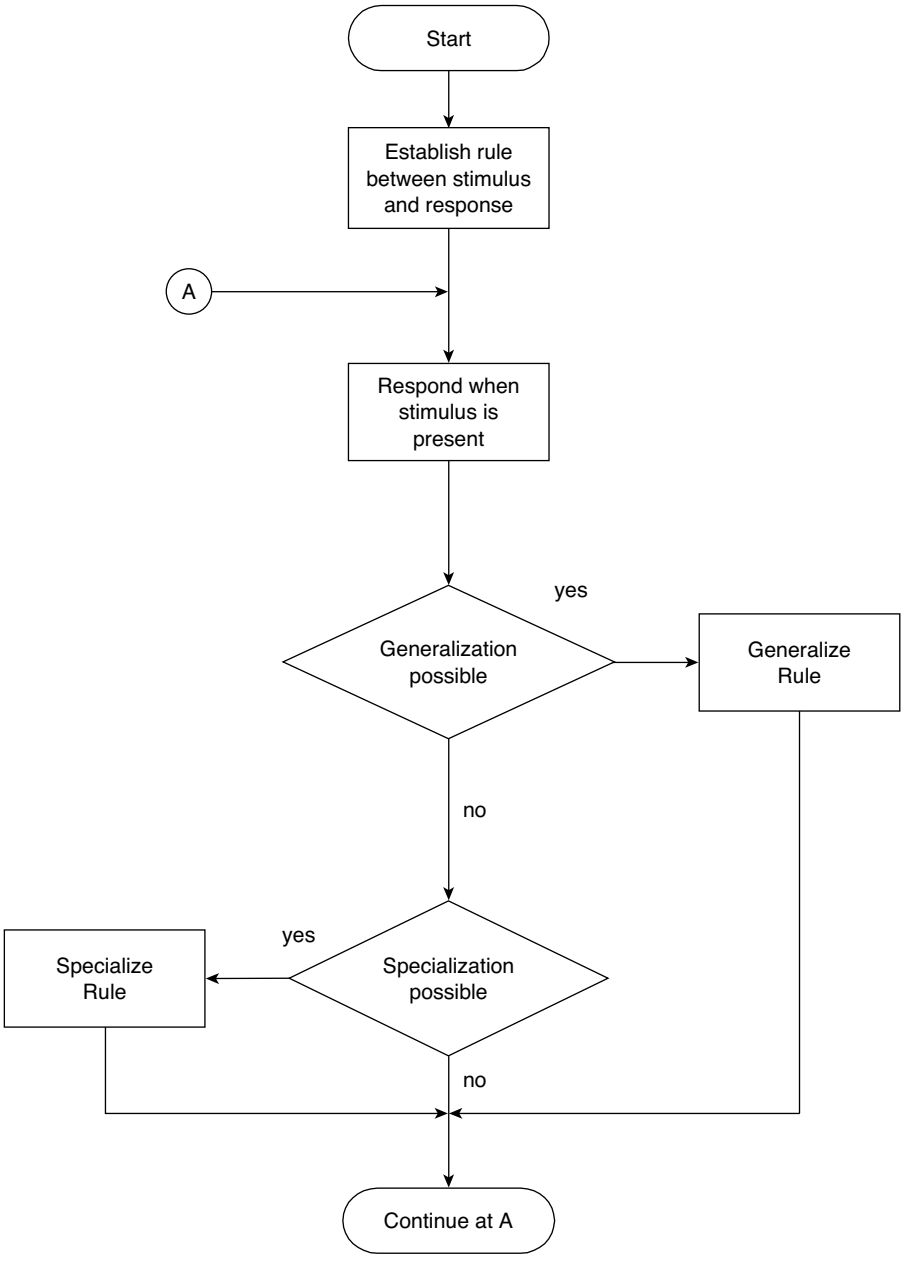


Figure 10.4 A machine-based algorithm for skill learning. Rectangles represent actions to be taken; diamond shapes represent decision options

The operation of this algorithm is exemplified by the following illustration: We build a highly advanced robot and provide it with the rule that "cats may be petted." Upon encountering a tiger, the robot is severely damaged. Once we have rebuilt the robot, the rule should be specialized to include a precondition, such as, "small cats may be petted"; the rule has been made more "complex" and less "general." As another example, recall the discussion in Chapter 9 (The Linguistic Approach) in which it was said that English language acquisition requires a refinement of the generalized "add-ed" rule for irregular verb past tense formation (e.g., "go-ed" becomes "went"). This learning model could also describe the study of linguistic cognition itself, whereby a theory evolves (e.g., the Wernicke-Geschwind model based on language deficits) and is subsequently refined or specialized on the basis of its failures (e.g., a lack of correlation between neurological foundations and hypothesized functionality). In this case a teacher would normally provide the reinforcement needed to correct the flawed behavior (e.g., a parent).

Reasoning is the ability to draw inferences appropriate to the facts and the situation. It is embodied in logical thinking and includes deductive, inductive, and abductive reasoning. A simple illustration will help to explain deductive logic. The machine is provided with the following predicates:

"Jane is either in the library or in class."

"Jane is not in class."

Computer programs are able to conclude that:

"Jane is in the library."

Even though computers have demonstrated considerable success in this regard, one would not conclude that they can "reason" based solely on their ability to draw correct inferences. Reasoning includes the ability to draw inferences that are relevant to the circumstances. One of the most difficult problems that confront AI scientists is the need to develop algorithms that can distinguish between the relevant and the irrelevant. Consider the circumstance in which an intelligent computer agent is called on to diagnose a disease or deficit on the basis of a given set of symptoms it has received as input. The computer may come up with a number of possible causes of the symptoms, many of which may be irrelevant. An experienced physician would know how to filter out some of the symptoms so as to zero in on a diagnosis. The computer may not have the ability to weigh the relative importance of each symptom.

Problem solving is yet another key element of intelligence. A machine may be presented with a photograph of an individual; can it identify the person in the picture? A robot is presented with a task; can it devise a strategy to complete the assignment? There are two problem-solving methods: special-purpose, and general-purpose. A **special-purpose solution** relies on the circumstances that pertain to the task. The **general-purpose solution** can be applied to a broad variety of problems.

One technique employed in AI is known as **means-ends analysis**. Consistent with this method, the machine assesses the current state of the system and chooses an action that will reduce the difference between the current state and the goal state. For example, a robot may be able to perform operations such as pickup, putdown, moveforward, and so on. From this set of operations it will pick that action that reduces the difference between the current state of the system and the goal state. This algorithm mirrors similar mechanisms that are used by human intelligence when human beings are faced with the same task that the robot has been given. The topic was raised in some detail in Chapter 5 (The Cognitive Approach II).

Perception is achieved when the intelligent entity, animal or machine, scans or samples the environment using the sensory equipment it has at its disposal, and the information is processed by internal processing mechanisms and converted to a set of abstractions that is made up of some combination of objects, features, and relationships. This can be difficult for a machine, as different views of the same object or scene must be correctly interpreted. Contemporary robotic machines are sufficiently advanced to the point where they can carry out the complicated tasks for which they've been designed (e.g., deliver mail in corporate offices, assemble parts, perform cleaning tasks, etc.) But such automata are far from maturation, as described in Chapter 12 ("The Robotics Approach").

Some argue that using language to communicate comprises the highest form of intelligence. It is relatively easy to write computer programs that enable machines to respond rather fluently, say, in English, to questions and statements. (See the discussion of Parry, above.) Responses may be fluid to the point where the computer's linguistic behavior is indistinguishable from that of a live human being. Can we conclude that the computer possesses any linguistic understanding? Conventional wisdom on this question would require that the intelligent entity, person or machine, had learned the language while it was a part of a linguistic community—had learned it interactively, with other users of the language. In the case of machines, this is not likely to happen until strong AI has been achieved.

AI Methodologies

Two methodologies, top-down and bottom-up, characterize AI research. **Top-down** approaches treat machine intelligence as a "high-level" phenomenon that does not depend on how the computer implements detailed or "low-level" operations. Such low-level operations would include such things as adding numbers or comparing numbers. In top-down methodology, we can ignore the neurological interconnections within the brain or, in the case of the computer, the electrical interconnections. The **bottom-up** model provides an alternative approach to the study of machine intelligence. Generally, this approach uses the neural network model that is used by the connectionist community. In this methodology, scientists explore the aspects of cognition that can be recreated via the use of neural nets. Figure 10.5 depicts the differing methodologies in AI research.

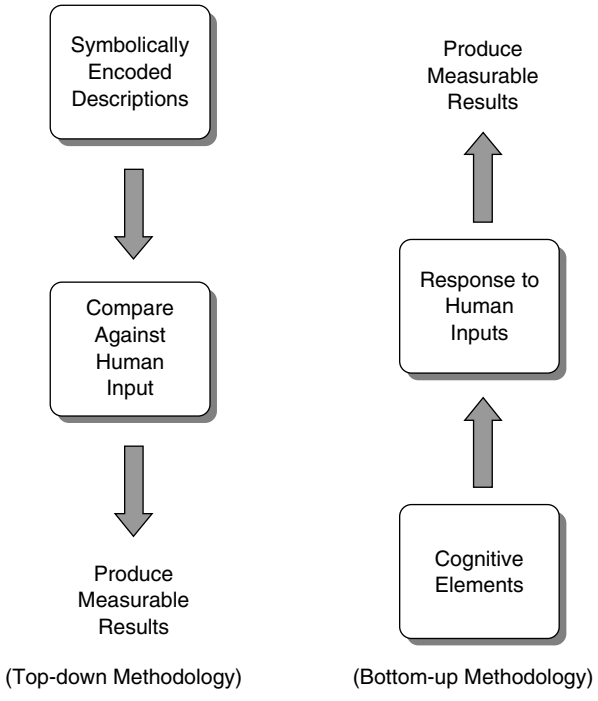


Figure 10.5 Top-down and bottom-up methodologies in AI research

To illustrate the difference between these two approaches: A machine has been programmed to recognize letters of the alphabet. The letters will be supplied to the machine in script form and presented to an electronically sensitive tablet similar to those found in modern handheld personal business notebooks. A top-down machine would include a database of the symbols and/or characteristics that correspond to each letter of the alphabet. As text is entered, a program within the device would compare this input against the database of parameters that define each of the letters. The device would make comparisons between the two on the basis of such things as the relative angles of the strokes, the lengths of the lines, and the intersections of the segments. It would then select the letter whose symbols and/or characteristics best match the input. It has made its decision on the basis of abstract considerations.

A bottom-up automaton would be wired in a way that resembled the neural configuration of the retina. The network would have 26 outputs (or more), depending on the number of characters it is to recognize. The pattern that is generated on the input tablet is presented to the machine (the program). As the machine was receiving the input, the output that responded more vigorously than the other 25 would become the appropriate result. In short, a bottom-up methodology would depend on "neural activity," while a top-down system would depend on descriptions, lists, or a database of knowledge that in some way defined the task.

The top-down methodology enjoyed early renown within the AI community. During the 1970s, Newell and Simon used the terms "symbolic AI" and "the Physical Symbol System Hypothesis" to describe the approach. Such terminology emphasizes the concept that the processing of structures (ordered arrangements of information) as well as symbols by a digital computer constitutes the essence of AI. Newell and Simon also maintained that the processing of symbols by the human brain is the basis of human intelligence. It is to be noted that in the decades since the Newell-Simon approach was at the height of its popularity, successes that have been owing to the use of the connectionist (bottom-up) model have tended to eclipse symbolic AI. However, as we shall see, Fuzzy Logic has recently provided new support for a (quasi-)top-down model. (See Chapter 11, Artificial Intelligence II: The Operational Perspective).

In summary, a top-down approach to AI is a problem solving approach that is based on the descriptions of the relevant features of the task, whereas the bottom-up approach depends on the neural activity of the intelligent agent.

The Computer as the Tool of AI Research

Whether by conscious design or not, the organization or **architecture** of the modern computer, especially the PC, parallels brain organization. (Or is it the reverse—have cognitive scientists modeled the brain and its functions on the basis of the computer model?) As noted in Chapter 1, cognitive science can be viewed as the study of the mind. Within the brain—the primary "seat of intelligence"—key elements of its organization include computation, memory, and communication. This model readily describes the modern digital computer. A functional view of the computer is shown in Figure 10.6.

The functional elements of the digital computer shown in Figure 10.6 include the Central Processing Unit, the Memory, the Input/Output, and the Communication Bus:

The **Central Processing Unit** or **CPU** executes the PC's set of instructions, and controls the order or sequence in which the individual instructions are executed. Performing calculations, logical operations, and data comparisons are fundamental capabilities of the computer. Calculations include operations such as adding, subtracting, multiplying, and dividing. Logical operations are employed by the computer to test the truth or falsity of a statement (e.g., "Is the patient's temperature exactly 98.6 degrees?"). Logical operations are important for determining the sequence of program steps. For example, if elapsed time (as determined by a clock) is greater than 10 seconds, then the next instruction to be implemented might inform a user that an answer to a

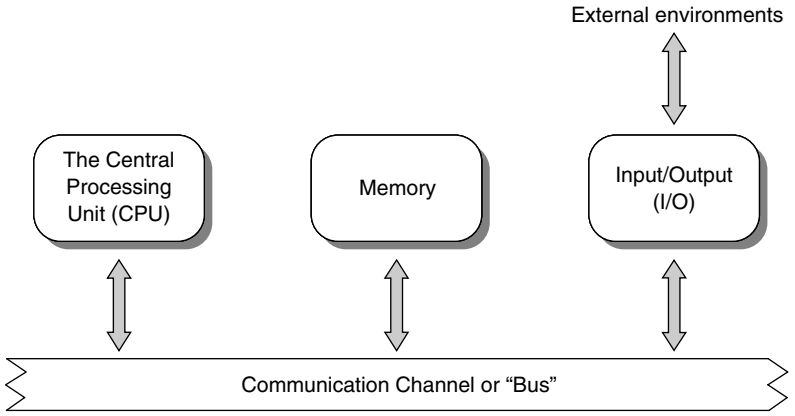


Figure 10.6 Functional diagram of a digital computer

question has not been received within the prescribed interval and would cause the computer to "jump" to another sequence of instructions that would address the failure to respond. The operative phrase here is "greater than," wherein a comparison between the actual time and the time at which the question was initially presented to the user is made. Although cognitive philosophers have, to various degrees, identified processes similar to those described above that take place in the brain, they have not identified a comprehensive structure within the brain that precisely implements the precisely defined properties of the CPU.

Memory functions include: storing the instructions to be executed, partial computational results, and final results. Memory is generally divided into **main memory** (short-term, volatile memory segments) and **long-term** memory (non-volatile elements, that is, the "disk" or "hard drive"). For example, the CPU may be instructed to update a series of patient records. As each record is modified, the **partial results** are retained in the PC's memory. Once all modifications have been completed, the new information may be permanently stored on the hard drive and a printed ("hard") copy of the results may be generated.

The **Input/Output** (**I/O**) provides all the electronic hardware and software resources that enable the computer to interact with the external world. "The external world" may refer to the human user, who interacts with the computer via an appropriate device, such as a microphone. Additionally, the PC can be connected to an external device such as another computer, or other type of electronic device (e.g., a digital camera).

The **Communication Bus** is the channel of communication among the aforementioned elements through which signals and data pass.

It is noted that the architecture represented in Figure 10.6 is a basic formulation of the computer. A number of alternative arrangements are possible, such as arrangements that include multiple processors, each having its own CPU, Memory, and I/O facilities. Such arrangements can greatly increase the speed at which programs are executed. The rate at which tasks or "jobs" are completed is referred to as the computer's **throughput**. Such architectures will not be discussed.

Evaluation of the Computer as a Model of Brain Organization

Some physiological psychologists maintain that the functioning of the body is too complicated to be supported by a "single computer." According to these scientists, the analogy of the brain as the repository of a "CPU" must

inevitably fail, as such a living CPU could not support the inordinate amount of neurological information processing that must go on if homeostasis (i.e., equilibrium) is to be maintained. Dworkin (1993) suggests a distributed model. A simple example of information processing that goes on in the absence of direct brain involvement is seen in human reflexes—for example, when we reflexively remove our hand from a very hot surface. Additionally, the connectionist approach (see Chapter 7, The Network Approach) considers processing and knowledge representation to be distributed.

Programming

The series of instructions that a computer programmer devises that enables the computer to run is called a **program**. One functional element of the computer that is not depicted in Figure 10.6 is the computer program, or software, that has usually already been pre-installed on your PC. This program allows all the functional elements of the computer to operate in a smooth or cooperative manner; it is referred to as the PC's **Operating System**. Originally, programs had to be entered into the machine as series of numbers. A computer cannot "understand" textual data; it can only understand numbers, and binary numbers in particular. Entering numbers proved to be time-consuming, and vulnerable to the introduction of error, especially for programs of any length. Numerical programs were replaced by those made up of mnemonic equivalents for numbers, and this proved to increase the efficiency with which programs could be written and installed on the computer. These mnemonic representations were called **assembly language** programs. And still later, programs made up of "English-like" statements, closer to those of natural languages, replaced assembly programs. Thus, to add two numbers a programmer could write the following instruction:

$$C = A + B$$

The computer is instructed to "go to" the location that stores quantity A, add to it the quantity that has been residing in location B, and store the result in a location named C. Languages built on statements such as this are called **High Level Languages** (HLL), because of their resemblance to natural languages. Modern programming has evolved in the direction of visual programming, in which icons or symbols replace operational notation. As computer capabilities approach those of humans, it may soon be possible to develop programs using spoken natural language commands. A programmer would issue

a spoken command to the computer to carry out an instruction (e.g., "add A to B" or "please add the quantity A to the quantity B"). AI would enable the computer to interpret such instructions. The highly abstract nature of AI that uses top-down methodological tools provides us with a means for attacking the problems that attend natural language programming.

Evaluation of Programming Languages

Early on, investigators who were trying to develop AI programs noted that the HLLs that were available to them were cumbersome. The software structures and facilities available in the existing HLLs did not lend themselves to the problems they were facing. Newell, Simon, and Shaw found it difficult to write instructions, in HLLs that were available to them, for an application that they were developing (i.e., the General Problem Solver, or GPS, that is discussed in detail in Chapter 5, The Cognitive Approach II). GPS uses lists of conditions to describe the current state of the world as well as the goals to be achieved. A process known as means-ends analysis is used to reduce the difference between the existing state of a system and the goal state. Newell, Simon, and Shaw subsequently developed the Information Processing Language (IPL) that was better suited for AI programming. An historical perspective of AI programming languages is provided in Table 10.2.

Application-oriented software packages that support AI activities such as Fuzzy Logic (MatLab tool) and Neural Net (Brainmaker) are currently available for research and development.

Alan Turing and the Great Debate

In Turing's seminal paper of 1950, he recognizes the difficulty inherent in attempting to define "thinking" and cleverly reformulates the task into an "imitation game" (IG).[1] As originally conceived, the IG included three people: an "interrogator" of either sex, a man, and a woman. This reformulation of the original question "what is thinking?" in the form of the IG has become known as the **Turing Test** (TT). Although more recent versions of the test have come into being, its original formulation sheds light on the culture that helped to spawn it, such as it was at the time of the formulation. For example, note the placement of the word "woman" in the following description:

> The new form of the problem can be described in terms of a game which we call the "imitation game." It is played with three people, a man (A),

Table 10.2 AI Programming Languages.

Date	*Program*
1956	Newell, Shaw, Simon (RAND, Carnegie Institute of Technology) introduce the Information Processing Language (IPL). Features of the program include: lists, associations, schemas (frames), dynamic memory allocation, data typing, recursion, associative retrieval, function arguments, streams, multitasking. Newell specified the language syntax, Shaw acted as system programmer, and Simon was the application programmer/user. IPL pioneered the concept of list processing. The first application was to demonstrate that the theorems in *Principia Mathematica* (Russell, Whitehead) could be proven by machine. While the recent version of IPL still runs today on readily available PCs, it has generally bean replaced by Lisp, with similar features but simpler syntax and automatic garbage collection.
1960	John McCarthy combines elements of IPL with elements of lambda calculus to produce LISP (LISt Processor). The lambda calculus was devised by Alonzo Church (Princeton logician) while he was studying the decision problem for predicate logic. (This was the same issue being studied by Turing when he described the Turing Machine.) LISP remains a fashionable language among AI investigators.
1973	PROLOG (from PROgrammation en LOGique) is conceived by Alain Colmerauer (University of Marseilles) and subsequently developed by Robert Kowalski (Edinburgh University). It implements a theorem-proving technique known as "resolution." This technique originates in 1963 (Alan Robinson, Argonne National Laboratory, Illinois). (PROLOG and LISP are used extensively in the Natural Language Computing Project at NYU.) PROLOG can determine whether or not a given statement follows logically from other given statements. For example, given "All logicians are rational" and "Robinson is a logician," a PROLOG program would answer "Robinson is rational." PROLOG is used in Tokyo (Institute for New Generation Computer Technology). Computers employing such programming elements are referred to as belonging to the "Fifth Generation" technology.
1990s	Application-oriented software packages that support newer AI technologies and research interests such as Fuzzy Logic (MatLab tool) and Neural Net (Brainmaker) are extensively used in research.

a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two, and cannot hear their voices. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by

> labels X and Y, and at the end of the game he says either "X is A and Y is B" or "X is B and Y is A." The interrogator is allowed to put questions to A and B. [Italics added.]
>
> The task of the man is to convince the interrogator that he is the woman. The woman, on the other hand, should follow a strategy of helping the interrogator and of answering questions honestly. Thus, she might add qualifiers to her answers such as "I am the woman, don't listen to him." A simple exchange might have the following narrative:
>
> Interrogator to X: Please tell me the length of his or her hair.
>
> X replies [when X is actually a man]: My hair is shingled, and the longest strands are about nine inches long.
>
> [X, the man, is attempting to mislead the interrogator.]

Turing takes this a step further when he asks the question: "What will happen when a machine takes the part of A (the man) in this game?"

The passionate debate that sprang up in the wake of Turing's paper has led to the current representation of the TT, which is generally agreed to have the following formulation. The test consists of: a human interrogator, one human responder, and a second "person"—actually a machine. The participants are separated by screens or partitions: they cannot see one another. This form of "neutral" communication precludes the introduction of clues such as the sounds of the voice and/or the idiosyncrasies of a participant's handwriting, as Turing had stipulated. The task of the interrogator is to find out which of the two candidates is the machine and which is the human being. The interrogator is to arrive at this conclusion solely on the basis of the answers he or she receives to his or her questions. If the machine can "fool" the interrogator, it is defined as "intelligent." Some people have described this version of the test as the ultimate "imitation game" (IG). The sketch in Figure 10.7 depicts the TT.

As per the test conditions, interactions between participants are expressed in natural language—an insurmountable obstacle for machine intelligence such as it existed at the time the test was proposed (as well as today, as the discussion in Chapter 9 [The Linguistic Approach] attests to). At the end of a specified amount of time, if the human interrogator cannot tell which of the candidates is the machine and which is human, the machine would be said to have passed the test and would be considered "intelligent." To understand the difficulty inherent in interpreting test results, consider that the interrogator could simply make a guess at the end of each trial and be correct in his or her judgments 50% of the time. Thus the machine would need to confound the

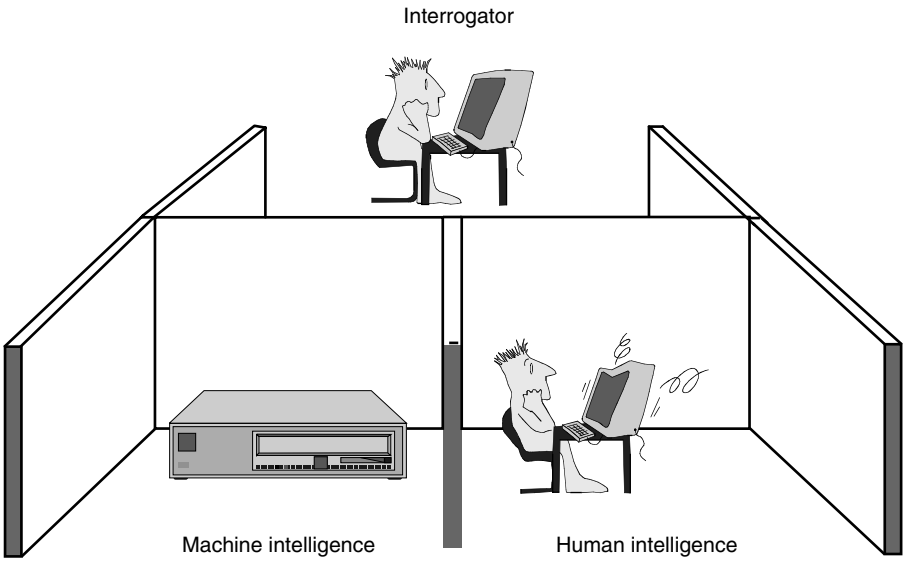


Figure 10.7 A sketch of the Turing Test

interrogator considerably in order to compensate for the biasing of the results in the interrogator's favor that is owing to the fact that correct answers may be produced by guessing.

The IG as depicted in Figure 10.7 is but a single instance of this type of game. For example, the game could be played according to its original format (i.e., the players are the interrogator, a man, and a woman). The objective of the interrogator is to determine which of the two participants is the woman (or man), whereas the objective of one of the participants is to confuse the interrogator into thinking that the sexes of the two participants have been reversed. It is now generally understood that the TT tries to assess if machine intelligence can imitate a human being, rather than imitate a woman specifically. (Did Turing believe that women could be imitated by machines but that men could not? But, this is something of a digression.)

Turing sought to differentiate between the "physical" and "intellectual" capacities of humans. Covering an intelligent machine with artificial flesh would provide no advantage. Moreover, this kind of action was not feasible when Turing was developing his thinking. Turing also proposed that the dialog take the form of a series of questions and responses, as it provided the opportunity for introducing any topic from the human experience. A typical fragment (from his paper) has the following form:

Question:	Please write me a sonnet on the subject of the Forth Bridge.
Answer:	Count me out on this one. I never could write poetry.
Question:	Add 34957 to 70764.
Answer:	(Pause about 30 seconds and then give an answer) 105621.
Question:	Do you play chess?
Answer:	Yes.
Question:	I have K (king) at my K1 (location of the king) and no other pieces. You have only K at K6 and R (rook) at R1. It is your move. What do you play?
Answer:	(After a pause of 15 seconds) R–R8, (check) mate.

If the machine were designed satisfactorily, then it could easily eliminate the strong objection that it could not imitate human traits such as slowness and inaccuracy; Turing thus rejects this argument. Turing limits the "machines" that may participate to "digital computers." He recognizes that the computers that existed in 1950 might not fare well in the IG, but envisions that "there are imaginable computers which could do well."

Turing's paper, the IG, and his ideas in general have been discussed, attacked, and defended since their inception. Within the cognitive and philosophic communities, the attitudes have ranged from profound admiration (the TT represents the birth of AI and passing the IG is the ultimate goal of AI) to extreme condemnation (the TT and the IG are useless and even detrimental to the goal of achieving machine intelligence). Turing's paper has given rise to arguments that center on consciousness (e.g., what is it?), behaviorism, and intelligence (e.g., what are its preconditions and prerequisites?). It has given rise to a contest: the awarding of the Loebner Prize, a gold medal and $100,000 that goes to the chatbot that passes the TT. (A **chatbot** is a program designed to generate text that is imitative of human conversation.) Each year a competition awards $2,000 (and a bronze medal) to the program that seems most human. The passion of the overall debate may spring from the recognition that a machine capable of humanlike performance would convince us that humans may indeed be nothing more than "machines"—a conclusion that goes against millennia of cultural tradition.

Turing made predictions that reflected his belief that a "thinking machine" was possible:

- In about 50 years [that would have meant around the year 2000], it will be possible to program computers that have a storage capacity of about 10^9 memory locations in which information (predicates and rules) can be retained, and this will permit them to play the imitation game so well that

an average interrogator will have not greater than 70% odds of making the correct identification after five minutes of questioning.
- At the end of the century [also around the year 2000], the use of phrases such as "thinking machine" will have become almost commonplace and will be used without the user's expectation of immediate rejection.

Evaluation of the Turing Test (TT) and Turing's Detractors

Turing anticipated arguments against his IG and persevered in his belief in the inevitability of a successful "thinking machine." He was considerate of the arguments that were opposed to his thinking. Table 10.3 summarizes these arguments and Turing's rebuttals.

The table lists arguments (which fall into eight categories) that summarize the various opinions that have emerged since Turing's paper first appeared. Turing's position can be viewed from a few perspectives: religious (Theological), cognitive (Heads in the Sand, Consciousness, Extra-sensory perception), behavioral (Disabilities), and Scientific (Nervous system continuity, Mathematical). Each of the arguments is summarized and Turing's rebuttals are given. He anticipated the arguments of potential detractors and rebutted those arguments.

Battle Lines: The Future of the TT

The TT has remained an icon of AI ever since its introduction in 1950. While many scientists and engineers have chosen to focus their efforts in AI-related areas and have been able to demonstrate numerous achievements in AI (as we shall see in the next chapter), success in the IG remains the "Holy Grail" of AI. Arguments (and counter-arguments) over TT have raged over the past half-century and we provide a brief overview of them (Saygin, Cicekli, and Akman, 2000).

The TT as an Operational Definition of AI

Turing proposed the TT as a "definition" of machine intelligence. Consider, for example, Searle's Chinese Room argument (see Chapter 2, "The Philosophical Approach"), which describes an IG scenario in which a set of input/output rules might appear to exhibit "intelligence"; however, this

Table 10.3 Arguments Against Turing's Proposals

Opinion	*Argument*	*Turing Rebuttal*
Theological	Thinking is a function of man's (God-given) immortal soul. No animal or machine can think.	This argument is a serious restriction of the omnipotence of the Almighty. He has the freedom to confer a soul (as well as a brain befitting thinking) on an elephant if He sees fit.
Heads in the Sand	The consequences of machines thinking would be too dreadful.	This is the most extreme formulation of this type of argument. If Man is superior to the rest of creation then a commanding position (over machines) will be sustained. Most likely intellectual people feel threatened by machines because they value the power of thinking more highly than others. Turing suggests that consolation is more appropriate to such thinkers and that believers might find solace in the transmigration of souls.
Mathematical	Gödel, Church, Kleene, Rossen, and even Turing himself have developed mathematical logic results demonstrating the limitations of discrete-state machines (i.e. computers). (In any sufficiently powerful logic system, there are some theorems that can neither be proved nor disproved.) Hence, the kind of machine that Turing proposes is not possible.	While these mathematical results are valid, this argument implies (without proof) that no such limitations apply to the human intellect. Man might be cleverer than any given machine, but then again there might be other machines cleverer again (than man), and so on.
Consciousness	No machine can write a sonnet, compose a concerto from thoughts and emotions (or know that it created such works), or experience grief, flattery, be made miserable by mistakes, be charmed by sex, or be angry or depressed.	According to the most extreme form of this view, the only way by which one could be sure that a machine thinks is to *be* the machine and to feel oneself thinking. The only way to know what a man thinks is to be that particular man; it is the solipsist point of view. "A" is liable to believe "A thinks but B does not," while "B" believes "B thinks but A does not." Instead of arguing continually over

Opinion	*Argument*	*Turing Rebuttal*
		this point, it is common to follow the polite convention that everyone thinks. (In other words, how do you define "thinking" and how do we know when it is achieved?)
Disabilities	Machines can do what you say but you will never be able to make one do "X" (where X might refer to kindness, resourcefulness, beauty, friendliness, having initiative, having a sense of humor, telling right from wrong, and so on).	Man continuously sees machines of limited purpose and therefore cannot conceive of a machine with greater capabilities. Many limitations ("disabilities") are associated with the very small capacity of most machines (of the time). Some of the limitations must be considered to be frivolous. For example, a machine might be made to "enjoy" strawberries and cream, but any attempt to make one do so would be idiotic. (How does one define the "friendliness" between a man and a machine, as compared to friendliness between a white man and a white man, or between a black man and another black man?)
Lady Lovelace	Babbage's biographer noted that the Analytical Engine had no pretensions to *originate* anything. It could only do "*whatever we know* how to *order it* to perform" (her italics). A variant of this objection: a machine can never do anything really new.	This can be parried with the statement, "There is nothing new under the sun." Who can be certain that "original work" is not simply the growth of the seed planted by teaching, or the effect of following well-known general principles?
Nervous system continuity	The nervous system is not a discrete-state machine. A small error in information can make a large difference in output. Therefore, a machine cannot mimic nervous system behavior.	Within the context of the IG, the interrogator will not be able to take any advantage of this difference (the resolution of the analog brain). A digital computer (within the IG) could be programmed to produce results indicative of a continuous organization (an organism).
Extra-sensory perception.	Telepathy, clairvoyance, precognition, and psycho-kinesis cannot be replicated by machine.	Statistical evidence for such phenomena is, at the very least, not convincing. Many scientific theories remain workable in practice in spite of clashing with extra-sensory phenomena.

designation of intelligence is at variance with concepts espoused by many in the cognitive community. Aunt Bubbles (Block, 1995) provides an attack on the TT as a behaviorist approach to intelligence. (Aunt Bubbles is an example of a dumb machine that would appear to pass the TT. The interrogator enters some statement or question. The computer searches its database and produces a response that the programmers have deemed appropriate. The interrogator now enters a second question and the computer produces a new response that is consistent with the second question and the computer's first response. The machine can do well with regard to the TT, but it is basically considered to have the intelligence of a juke-box.) The arguments in support of, and in opposition to, the TT are basically philosophical. The reason for this is to be found in the following argument: if a machine passes the TT, then by definition it should be labeled as "intelligent"; if a machine fails the test, we cannot say for certain whether it does, or does not, think in an intelligent manner. We can agree, however, that the TT tests for "human-like" intelligence. We might use the TT as a means for gathering inductive evidence about machine intelligence as well as information that will help to resolve the cognitive mind-body issue. It remains to be seen whether a realistic and objective method of assessing the test regarding successful machine implementation will be developed any time soon.

The TT and Behaviorism

TT-like tests seem to be the best methodology we have for forming conclusions about machines' "minds," even though such tests may intuitively support a behaviorist viewpoint. We currently grant intelligence to humans on the basis of considerations of behaviors, so why not do so for the machine? Behavioral evidence is the best evidence we have for other minds. If, someday, our basic understanding of human intelligence becomes altered, then an alternative test could be developed for the machine. If a machine passes the TT some day, we will have to accept the fact of machine evolution in the same way that we have accepted animal evolution.

Is the TT Too Easy or Too Difficult?

Some researchers criticize the test as being too easy (Bringsjord, 1994), using the argument that it enables only the assessment of "verbal" intelligence. It is incumbent on those who promote this view to design an "acceptable" test, and

if the machine were to pass such a test, to make clear why the machine does not deserve to be called intelligent. If a machine were to pass the test because it had used, essentially, a "bag of tricks" (as in Searle's Chinese Room, for example), should we consider that the human mind is also a bag of tricks? After fifty years of trying, we have developed only rudimentary natural language processing systems (Weizenbaum, 1996). One example is the chatbot, which can carry on conversations with people via the Internet. None of the natural language processing systems passes the TT. Yet, other researchers declare that the TT is too difficult. The Internet alternative that can find information possesses a level of "behavioral" intelligence that would seem to approach the TT level. One of the main problems is our limited understanding of natural language processing and generation in humans. If it turns out to be the case that human natural language processing is fairly impervious to our analysis, it will be impossible to model such processing on computer functioning in a successful manner, if at all. However, deficiencies in the TT do not preclude machine intelligence. Modern computers now perform many tasks that we would not hesitate to say had required intelligence if the tasks had been executed by humans. (Some of these will be discussed in the next chapter, Artificial Intelligence II: Operational Perspective.) A natural language machine answering inquiries on a well-defined topic is clearly quite remarkable. We should not dismiss the machine as "lacking intelligence" if it fails the TT. A machine's passing the TT can be considered to be a sufficient condition of or proof of its intelligence; we should not dismiss practitioners of AI as not being able to build intelligent machines.

Overall Evaluation of the AI Concept: Summarizing the Meaning of AI

Alan Turing's predictions have failed to materialize in the fifty years since his treatise appeared. Most agree: his bold challenge has left us with an extremely difficult task. Initial enthusiasm has given way to pessimism regarding the ultimate goal—not just an intelligent agent that passes the TT, but one that possesses the capabilities of its human creator. Critics have expected too much too early. Perhaps the TT is the "Holy Grail," and, like the solution to Fermat's Last Theorem, its attainment may take centuries. (Fermat scribbled a mathematical theorem in one of his personal books, as well as a statement that he had a solution, but the solution would not fit in the margin of his book. The solution actually required some 200 pages of mathematics.). To create a successful TT, we will have to answer the following questions:

What is the language/cognition relationship?

Is language autonomous with respect to other cognitive abilities?

Can computers "understand" language?

What is a simulation?

Psychologists, computer scientists, engineers, philosophers, and linguists have wrestled with these questions and similar questions for many years. A successful TT may be a sort of culmination that follows on the heels of a number of other related accomplishments.

Many within the cognitive scientific community adhere to the belief that humans cannot, and should not, be viewed merely as machines. Yet, humans have an abiding fascination with the idea of building machines that recreate the entire human experience, including consciousness (even though it is a complex concept with no universally agreed-upon definition). Consider the contemporary interest in "cloning." And, as we were developing material for this text, the latest science fiction adventure movie made its appearance. The name of the movie: "I, Robot."

Although there are many definitions of AI, common themes or elements are evident in these competing definitions, which depend on the underlying interests of the researchers:

In engineering, AI is using knowledge to solve real-world problems.

In cognitive science, AI is a codification of knowledge that will finally explain intelligence.

Using the modern computer (PC), workers—no matter what their specializations—seek to reproduce (simulate) nothing less than human intelligence and mental processes.

The modern computer has the capacity to recreate "human-like" behavior for limited-environment problems. Alan Turing signaled a seminal test of machine intelligence in 1950. Some consider the TT to be detrimental to AI (Hayes and Ford, 1995; Whitby, 1996). AI's objective is to make computers perform "intelligent" tasks and thus make life easier. If that is the principal purpose of AI, then successful TT programs would not be terribly useful. The TT may be a game; our trying to develop computers that can communicate with humans using natural language may lead to seminal insights into how the human mind works. With very few exceptions, cognitive scientists as well as philosophers consider this a noble outcome.

In Depth: Behaviorism and Ned Block

One of the more vocal participants in the TT debate has been Ned Block (1981), who notes that the judges in the IG can be fooled. (Searle's Chinese Room argument is similar to Block's scenario.) "Mindless" machines given the TT rely on simple tricks to pass it. Block's hypothetical machine, which was designed to pass the TT, would have been designed in the following way (and such machines do not require sophisticated information processing algorithms):

- All possible conversations of some given length are stored in the computer's memory (database or knowledge base).
- At least one participant in the TT is "making sense" (is not irrational) in the natural conversations such machines can carry out.
- The set of conversations that can be carried out in a fixed amount of time—one of the parameters (constraints) of the test—is finite.
- Although an inordinate number of these conversations exist, they can be stored in the memory of the hypothetical machine, and then the game might proceed as follows:

  Judge: types in a character string; call this sequence A
  Machine: Finds a conversation that begins with string A and proceeds to generate a second sentence in the "conversation." Call the machine's response string B.
  Judge: Types in string C in response to B.
  Machine: Finds a conversation with ABC as its sequence and types out response D.

The game would proceed in this manner until the conversation is terminated, at which point the judge must decide if the machine is human (i.e., has exhibited human"intelligence"). In Block's view, the processing capability of such a machine is akin to that of a jukebox. Since he believes that the machine in the scenario will pass the test, he concludes that the test is an inadequate measure of intelligence. The problem, according to this line of debate, lies in the "behaviorist" approach inherent in the TT. Block therefore defines a concept which he names **psychologism**, wherein intelligent behavior depends on the character of the internal information processing that produces the observed results. What follows from this concept is depicted in Figure 10.8.

As shown in the figure, both systems produce the same behaviors, yet System 1 may be granted intelligence, whereas System 2 may be judged to be devoid of any intelligence by virtue of its different information processing

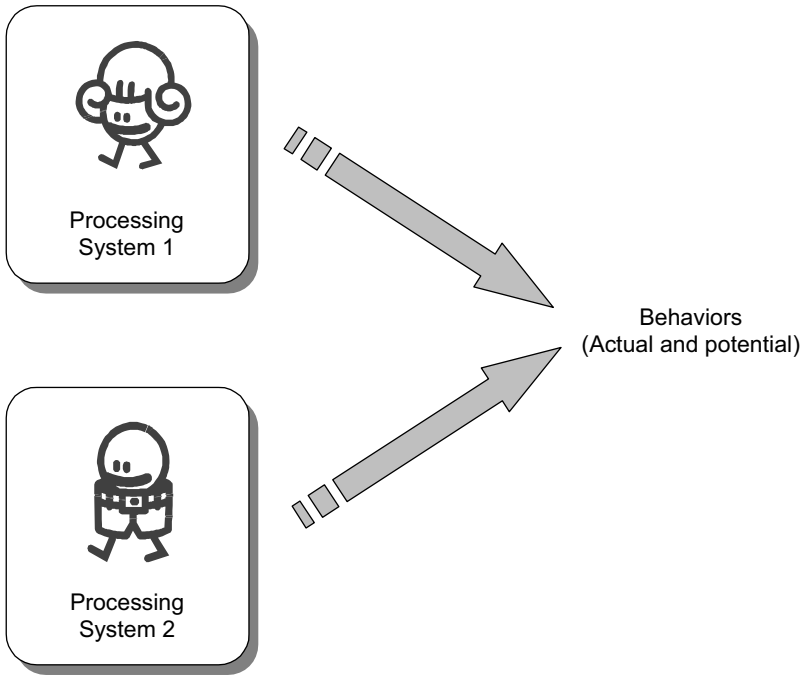


Figure 10.8 Ned Block's formulation of psychologism

structure. It is possible for both systems to exhibit intelligence, as might occur if we encountered "Martians" who could engage in creative activities. Block would not preclude Martian intelligence just because Martians are "different"—have unique processing systems. Regrettably, we as humans often assume a xenophobic attitude toward people of other cultures and are prone to deny them any "intelligence." Because an information processing mechanism is completely different from a human processing model does not mean that the "alien" entity lacks intelligence. Block is not critical of the TT in its own right. Rather, he is using the TT as a focus for his argument against a behaviorist approach to intelligence. Thus, he concludes that the embodiment of intelligence in tests such as the TT may lead one to find intelligence in an entity entirely lacking in such intelligence and to the possibility that we deny intelligence to a truly intelligent system. He concludes, "We should specify, in a *non-question-begging* way what it is for a sequence of responses to verbal stimuli to be a typical product of one or another style of intelligence." Block's formulation of the TT becomes:

> "Intelligence (or more accurately, conversational intelligence) is the disposition to produce a sensible sequence of verbal responses to a sequence of verbal stimuli, whatever they may be."

Good questions are not required; the system would have the capacity to emit sensible responses to anything that the interrogator *might* say, not just what *is* said.

Evaluating the Block Approach

The modified (neo-TT) conception of intelligence is still not sufficient to save it from a behaviorist label. The previously noted jukebox would still qualify as intelligent because it has the capacity to emit sensible verbal output to any verbal input. The intelligence that it exhibits belongs to the programmers, not to the machine itself. Machines that can learn and solve problems could qualify as being intelligent, because it could be argued that the "intelligence" belonged to the machine, not to the programmers. Consider an AI built on the Connectionist Model (described in Chapter 7) that *learns* to recognize patterns.

Block's argument is vulnerable in yet another way. Richardson (1982) doubts that Block's machine can imitate known conversational abilities. Humans understand and produce sentences that they have never heard or previously uttered. They can adapt to novel situations and maintain coherence during a discourse. The brain is not a repository of responses; it must be able to build a virtually unlimited set of sentences out of a finite list of words. If such a database of utterances is unlimited (over a lifetime), how can a team of humans gather these utterances and enter them into the computer's memory (in a finite amount of time)? According to Richardson, if list-searches satisfy the neo-TT, the test must be too weak. Given the present state of understanding of human cognition, based on current cognitive and linguistic research, it may be unlikely that Block's approach can succeed in modeling the human mind. There may come a time when Block's "sensible" strings of sentences can be enumerated and we will then judge intelligence on the basis of the neo-TT. However, on that day we would have all the psychologistic information that we require and we would no longer be interested in any sort of TT.

Block's argument is useful because it sheds light on the weakness of the behavioral approaches inherent in Turing's proposals. We may receive better insights as to how the mind works and therefore a better means to judge another entity's cognitive capacities. Until then, we have little that is more comprehensive than behavior on which we might base such judgments.

Minds On Exercise: Play the IG

Play the IG. Form a group of three: one will act as an interrogator, one will play the role of the male, and one the role of the female. To reproduce Turing's formulation, the team should communicate using one of the Internet "chat" tools or a software application such as Blackboard. The length of the period of interrogations should be agreed upon in advance. After the interrogations have taken place, the interrogator should decide which participant is playing the part of the male and which is playing the part of the female. Compile statistical results for all groups to determine what percentage of choices was accurate. Can you draw any conclusions from the results of this experiment? (If there is sufficient statistical data, the group might attempt to resolve a null hypothesis, such as, "The TT can measure intelligence." This would require some knowledge of statistical hypothesis testing.). The game should be structured in such a way that it is in accordance with the TT as it is viewed currently. For example, the group should set down the "rules" regarding what kinds of questions can be asked, such as by deciding what kinds of categories should be permitted (e.g., politics, history, weather, etc.).

Food for Thought: Discussion Questions

1. Could a computer replace one of the participants in the Minds On Exercise?
2. Explore the life of the atheistic physician, Julien Offroy de La Mettrie, one of the most radical proponents of the man-machine thesis of the eighteenth century.
3. Come up with examples, similar to the automatic pilot, of automata that do not "think" in the manner a human being thinks in the same or a similar domain. Look for other examples in the arts (e.g., plays) that deal with automata and their implications for machine intelligence.
4. Develop an example in inductive logic and evaluate the accuracy of the conclusion.
5. What's wrong (or not wrong) with the following argument?

"If each man had a definite set of rules of conduct by which he regulates his life, he would be no better than a machine.

There are no such rules.

Therefore, men cannot be machines."

As you discuss the logical problem, try to identify the mental processes you are using. Is this "intelligence"? Could a computer be programmed to produce similar results?

6. Does consciousness have to precede intelligence? Develop arguments for both affirmative and negative answers to this question.

7. Go to the website http://alicebot.org. A number of conversational computer programs have appeared since Weizenbaum first developed ELIZA. One of the more engaging and remarkable examples is ALICE. "A.L.I.C.E." is an "Artificial Linguistic Computer Entity" capable of intelligent conversation. If you were to interact with ALICE and did not know better, you might think that the program was really a human correspondent. The foundation supporting this enterprise, the ALICE A.I. foundation, is dedicated to the development and adoption of AIML (Artificial Intelligence Markup Language). On the website, locate the line "***Free Live Chat:*** with the award winning A.I. chat robot **A. L. I. C. E.**" By clicking on "**A. L. I. C. E.**," you will initiate an interactive conversation with ALICE. You can introduce yourself to ALICE (e.g., Hello, my name is John.), or ask ALICE any number of questions. See if ALICE knows what the word "algorithm" means. Try to determine what constitutes "cased-based reasoning." Ask ALICE about the concepts discussed in this text (e.g., connectionism). Can ALICE learn and/or remember what you tell it during the conversation? Do you think that ALICE is an intelligent agent?

Note

1. Alan Turing's seminal 1950 paper, *Computing machinery and intelligence*, should be read in its entirety; it is nothing less than a tour-de-force of issues in AI. The reference is readily found in several places on the Internet (www.abelard.org/turpap/ turpap.htm being one such site). The paper includes an intriguing description of a "human computer," which likens a cognitive model to the architecture of a digital computer.

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Asimov, I. (1950). *I, Robot*. Garden City, NY: Doubleday.

Boden, M.A. (1977). *Artificial intelligence and natural man*. New York: Basic Books.

Gelernter, D. (1994). *The muse in the machine: Computerizing the poetry of human thought*. New York: The Free Press of Macmillan, Inc.

Hogan, J.P. (1997). *Mind matters: Exploring the world of artificial intelligence*. New York: Ballantine.

McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. Natick, MA: A.K. Peters.

Simon, H. (1996). *Sciences of the artificial,* 3rd ed. Cambridge, MA: MIT Press.

11

Artificial Intelligence II: Operational Perspective

"Artificial intelligence is the science of how to get machines to do the things they do in the movies."

—Astro Teller

Introduction

As a direct consequence of human aspirations to be godlike, our history is replete with examples of attempts to create artificial beings or machines that imitate them (McCorduck, 1979). Long ago, Plato gave voice to the idea that our ability to be rational—to think logically—was closely tied to intelligence. This has been a driving force in AI.

From a practical point of view, AI is concerned with developing machines that solve real-world problems, particularly those requiring an intelligent agent (e.g., deciding the best travel route to take between two cities). The closely related disciplines of engineering and computer science have provided the hardware and software organizations that best implement AI within computer-based automata. AI has been defined in Chapter 10 and cognitive elements of intelligent agents (e.g., humans) have been discussed throughout the text. Underscoring this operational viewpoint are several well-defined subjects:

machine-based representation of knowledge (knowledge representation, or KR), formalized methods of logic and reasoning, and the associated topic of Expert Systems (ESs), Fuzzy Logic, and Artificial Neural Nets (ANNs). We begin by discussing the nature of the abstract representation of information in order to understand how such organizations accommodate to the strictures imposed by the computer. We then explore some reasoning methods that parallel human cognitive processes. KR and reasoning are integrated within ESs, which imitate some forms of human intelligence. After the limitations of ESs are considered, an alternative approach to the design of machine intelligence provided by Fuzzy Logic systems will be explored. ANNs will be briefly considered in light of the more comprehensive treatment previously provided in Chapter 7 (The Network Approach). Finally, an overall evaluation of such machines will summarize current issues that face those who undertake the design of intelligent agents with machine implementation.

The Practical World of Artificial Intelligence

The world that we inhabit has grown increasingly complex and to an increasing extent we have come to rely on machines to make efficient use of our material and human resources. While some express fear that the economic goal of AI is to replace human workers with machine equivalents, we should remain aware of some attractive prospects afforded to us by such machines. In some activities that humans believe are intelligent, (e.g., chess playing, computing math problems) machines may already outperform humans. However, the problems that weigh on replacing a human worker with a machine remain currently insurmountable—computers are not close to achieving the perceptive, finely-honed reasoning, and manipulative capabilities of adult humans. From an AI perspective, machines can currently demonstrate the intellect of a low order insect in some aspects of perception and information processing. Pursuing the study and development of machines capable of true AI has another positive outcome: the more we try to replicate human intelligence the more we may learn to appreciate and understand human intelligence.

Goals of Practical Machines

The criteria for success in machine-based AI can be summarized as follows:

- Does the application have a clearly defined task?
- Does the application solve a real problem?

- Are extensions to new opportunities possible?
- Does the solution embody a well-defined architecture or organization, or is the result impressive but ad hoc and suitable only in a limited number of cases?

A small sampling of the exciting practical opportunities within AI include: the mining of information (i.e., data identification and interpretation), resource allocation, uncovering relationships that exist between scientific variables or rules that describe those variables, developing effective control and problem solving strategies, improving the design of products that we use for activities of daily living, explaining decisions of all kinds, and identifying risks. These uses extend to the practice of medicine, commercial enterprise, law, social organization, and scientific investigation. The "intelligent computer" can be of enormous help to us.

However, designing computer-based machines that are intelligent is not the same as building computers that simulate intelligence.

We should not be obsessed with mimicking human intelligence. At the same time we should not reject methods that science tells us are involved in human intelligence. Here are some samples of the ways in which cognitive theories help to create useful machines:

Help experts to solve analytical problems in mathematics, theorem proving, and symbolic calculations, and in the study of the equations of complex object interactions. Cognitive models for problem solving are discussed in Chapter 5 (The Cognitive Approach II).

Help to design new devices. Find new ways to reduce the number of components in a design. (See, for example, Brooks, 2002.)

Help us to learn from examples. Experience-oriented learning provides an example of how to extract conclusions about new experiences using past knowledge. This parallels one aspect of human learning, and ANNs (see Chapter 7, The Network Approach) reflect mechanisms for drawing conclusions about an unknown stimulus. For situations involving greater abstraction, programs can examine databases to extract regularities. One such example is the ID3 system—discussed below—that examined thousands of databases and was able to produce identification rules in instances such as credit assessment and disease diagnosis (Quinlan, 1986).

Help to answer questions using structured or free format. START (Katz, 1997) is a system that enabled journalists and scientists to make inquiries

about the Voyager spacecraft, its journey to Neptune, and the solar system. One such question (and the answer) is noted:

Query:	What color is Neptune?
START response:	When you look at a photograph of Neptune, you see the planet's atmosphere and the tops of clouds. Neptune is bluish because the methane in its atmosphere reflects the blue and green frequencies of light. The methane absorbs red and yellow frequencies of sunlight.

Approaches to the Design of Intelligent Agents

Figure 11.1 depicts various approaches to the design of intelligent machine-based agents and they parallel cognitive models that are explored in other chapters.

The "top-down" approach in the figure reflects the idea that we can build a machine that "imitates" our concept of how human beings think at their most abstract level. It is a machine-based cousin of cognitive functionalism (see Chapter 2). This approach has given rise to ESs and historically these machines were the first to achieve practical success.

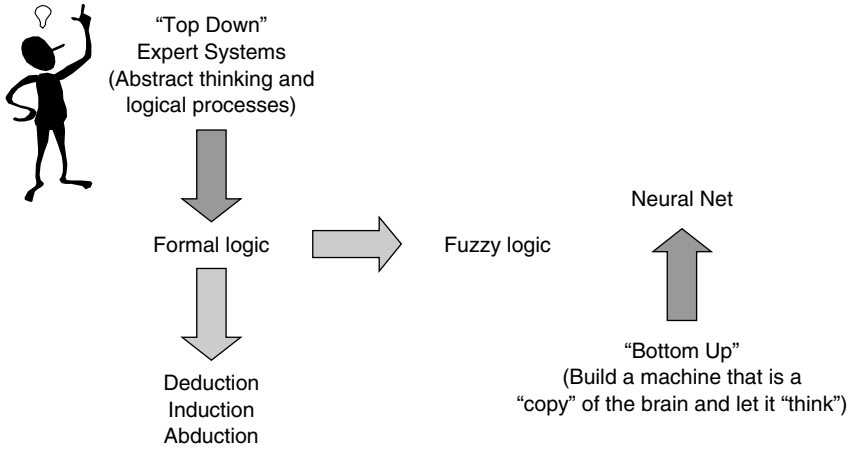


Figure 11.1 Diagram of operational elements of AI

The "bottom-up" approach uses the idea that we might literally build a neuronal model of the brain and see if we can train it to think and make decisions as humans do. This is a direct outgrowth of research in the neurosciences (see Chapter 6) as well as in the area of cognitive networks (see Chapter 7).

In this chapter we will concentrate on the basic concepts of KR, machine logic, ESs, and Fuzzy Logic as machine implementations of cognitive concepts. Comprehensive technical descriptions of ESs and Fuzzy Logic can be found in numerous texts, of which the Principe and Barr texts are representative (Principe et al., 2000; Barr et al., 1982).

Machine Intelligence, Knowledge, and Machine Reasoning

One of the goals of AI is to create a machine that performs computations modeled on human mental processes (i.e., an ES). In order to appreciate an ES, we need describe two important elements of such systems: KR and machine reasoning. A typical session with such a machine includes the following steps:

- A statement or question is submitted;
- The machine translates this into an internal representation or abstraction with minimal—ideally no—redundancy;
- Using an internal representation of the fact or question, relational statements (production rules), and the rules of logic, the machine arrives at results;
- The machine translates the results (e.g., an answer or explanation) from internal representations to forms (e.g., language, visual, auditory) that can be readily interpreted by a human user.

Machine Representation of Knowledge

The Cyc Project

To give you a sense of the scope and complexity of the task of attempting to represent knowledge within a machine, we describe one example of an ambitious attempt to do so. The name "Cyc" (for what is likely the largest experiment in symbolic AI) comes from the word "encyclopedia." In 1984, Doug Lenat, at the Microelectronics and Computer Technology Corporation in Texas, started a project in which he undertook to assemble a comprehensive database of everyday common-sense knowledge. The knowledge base (KB) is now the property of Cycorp, Inc. If it achieves its objectives, the KB will serve

as a foundation for ESs. Such systems equipped with "common sense" might be able to achieve a high level of performance and would be less prone to errors where "common sense" is lacking. For example, in one such ES, the program might attempt to diagnose a bacterial cause for symptoms stemming from a gunshot wound in which the patient was bleeding to death. (The system might draw this inappropriate conclusion because the user failed to include the fact that the patient had suffered an entry wound in his/her abdomen.) In another instance, the ES might prescribe an absurd drug dosage when the patient's weight and age were accidentally interchanged.

In the AI community, "common sense" refers to the very large body of worldly knowledge that humans use in their daily lives. Describing simple activities requires a mass of knowledge to convey an accurate representation of our environment. Consider some examples of common activities: one can pull an object with a string but cannot normally push an object with a string; an object resting on another object being moved usually moves with the object; water flows downhill; people who live in cities usually do not go outdoors when undressed; and cause precedes effect. Computers intended to reason must avail themselves of millions of facts. As one investigator, Terry Winograd of MIT, notes, "It has long been recognized that it is much easier to write a program to carry out abstruse formal operations than to capture the common sense of a dog."

Basically, the Cyc project requires data entry for millions of assertions—a formidable task in itself. During its first six years of existence, Cyc built a knowledge base (KB) of one million assertions. Lenat has estimated that it will take 100 years for two people entering such assertions to accumulate 100 million assertions. This is estimated to be the magnitude of the KB that would need to be assembled before Cyc might be truly useful. At any given moment, one may encounter 30 Cyc personnel entering data—they are referred to as "cyclists." These people examine newspapers, magazines, advertisements, encyclopedia articles, and other source material and extract information that is subsumed in the various publications. What Lenat had in mind was a KB that people or machines would have to have if they were to be able to understand an encyclopedia. One problem immediately comes to mind for this mechanism: Can we trust the cyclists to truly understand or interpret the underlying prerequisites that they are studying?

Cyc uses its common sense knowledge to draw inferences. For example, a statement such as "John is wet" could be deduced from the fact that "John has completed running the marathon" and three other facts: that running in such an event requires considerable exertion, that people sweat after great exertion, and that sweating produces a wet result. Typical of the entries are such things as "Every tree is a plant" and "Plants eventually die." In response to a query

such as "Will trees die?," the inference engine could conclude the obvious and respond correctly. Cyc has its own language, which is based on concepts related to computation with predicates and a syntax that is compatible with one of the so-called "AI languages" (Lisp). Knowledge engineering remains a focus of work being carried out in conjunction with Cyc, with particular attention being given to providing the Cyc system with the ability to communicate with its users in "natural language" as well as expediting the knowledge formation process.

Evaluation of the Cyc Project

The goal of trying to represent large amounts of knowledge leads to difficulties for Cyc. For example, Cyc has trouble in automatically searching for information that is "relevant" to a given problem. For symbolic reasoning as embodied in AI, the process of updating, searching, and manipulating a large structure of symbols in realistic time frames, as envisioned in Cyc, is not feasible. The underlying problem is one of scaling small databases into those large enough for genuinely intelligent systems. Critics believe that Cyc will succumb to this problem before it will be able to achieve human levels of knowledge. Other problems that confront Cyc are those related to knowledge representation. For example: how are basic concepts such as those that have to do with causation to be analyzed and represented within the KB? At present a significant part of the development effort is devoted to finding ways to search, update, reason, and learn (and particularly to learn through analogizing). Some maintain that this is the wrong approach—it "puts the cart before the horse."

Objectives of Knowledge Representation (KR)

Internal representation or abstractions of information are key elements of the interaction described above. The machine translates the natural language formulation of input from a user into a stylized, virtual equivalent. When the machine has developed a response, it must translate the abstract information into some natural format (e.g., auditory). Information within such machines consists of stylized or abstracted versions of the world. Such descriptions are called **representations** and were formally introduced in Chapter 1.

With respect to machine representation of knowledge, practitioners within AI concern themselves with questions such as:

- What is knowledge?
- How is knowledge to be represented?

- What is the relationship between mind, consciousness, and intelligence?
- What is the structure of language, and how does it relate to the reconstruction or initiation of thoughts?

Such questions have been addressed in other chapters in this work. In particular, the cognitive aspects of information representation were discussed in the main in Chapter 1 (categories of knowledge representation), Chapter 2 (the knowledge acquisition problem), and Chapter 4 (Neisser and cognitive psychology). In this chapter we will examine how information may be represented within a computer.

Early attempts to represent knowledge paralleled the way that it is represented in our cognitive centers. This is a problem, as we are still trying to fully understand how knowledge is represented in the human mind. One approach is to consider that intelligence is accessible via searches for heuristic information. (Note the discussion of the Cyc Project, above.) That is, we can resolve a problem or question by searching for the appropriate facts with the help of some general guidelines. As Marvin Minsky noted in 1982, "Everyone knows that if you try enough different things at random, eventually you can do anything." This would require extraordinary amounts of information and, in addition, it would take too much time to solve even relatively simple problems (Minsky, 1982). Currently, we recognize two kinds of knowledge: domain-specific knowledge, such as information within medicine, finance, or psychology, and general background knowledge. We have been able to provide computers with domain-specific knowledge; but as yet, the problem of organizing general background knowledge does not have a satisfactory solution because this kind of knowledge cannot be easily arranged in a systematic way as compared, say, to statements about a disease in medicine.

Knowledge representation is a central challenge for AI and is far from a trivial problem. (See Chapter 1 for a discussion of the cognitive dimensions of this problem.) Organization of knowledge within machine reasoning systems must address a range of issues, including how to store and manipulate information in a formal way so that it can be used by mechanical reasoning systems as well as be accessible to human users.

Characteristics of KR

KR provides us with a way to view objects in the world. Our representation impacts the ways in which we can draw inferences (Davis et al., 1993). Practically speaking, several KR organizations can be found within the modern ES. Whatever the specific architecture, these models address the roles that KR plays.

First, and foremost, a KR is a substitute or **surrogate** for the object or idea it seeks to symbolize. It should enable a computer to draw inferences via reasoning or thinking rather than acting. For example, if we were developing an AI to design bicycles, we would need to describe objects such as wheels, chains, handle bars, and so on. We would also need to put into it concepts such as the action ("process") that a wheel can take. Our descriptions are ghostlike or virtual images of these objects that only truly exist in the real world. In this view, reasoning itself is a surrogate for action in the world. A good surrogate must consider two things:

- What is it to be used for? There must be a correspondence between the surrogate and its equivalent in the real world.
- How good is the surrogate? Perfect correspondence is not possible. For example, the object in the real world may have a location that is not (and should not be) specified in our representation.

It should be noted that because surrogates are, of necessity, flawed, the inferences are inevitably flawed. However, a good representation can minimize (or even eliminate) error for a specific intended task.

In addition, a KR answers the question: "How shall we think about the world?" In other words, it represents a set of ontological features. **Ontology** is that part of philosophy (metaphysics) that is concerned with the nature of being, reality, or existence. Representations are views of the world. Like human vision, they may clearly focus on specific relevant elements, but this is accomplished at the expense of other parts becoming blurred—humans have reduced peripheral vision. In engineering, we represent the world in terms of very specific components. However, this is often accomplished without reference to the dynamic interaction of these components. On the other hand, medical diagnosis is viewed in terms of rules (see MYCIN, below) and these rules are different from the rules used in other medical ESs (e.g., the frames approach taken in INTERNIST; frames use prototypical descriptions—as opposed to the rules of MYCIN—and are described below).

A KR will include an **inherent fragment of intelligent reasoning**. A specific entry can include the set of inferences or operations that are sanctioned. For example, with an appropriate representation, we may imply or conclude that an ape "is an" animal. (An example of a hierarchical **isa** structure is shown in Figure 7.9.) Using descriptors such as age, weight, gender, and spouse, we may infer that man is a person or alternatively that our pastor is a person. This role that a KR plays follows from our conception that the representation is intended to provide an insight into how people reason. The reasoning fragment is implicit in the way we structure the KR technologies.

A KR is a medium for **efficient computation**. If we are to use a representation in any feasible way, we should be able to compute with it. Therefore, computational efficiency is strongly tied to KR. For example, we can include several kinds of information within our structure, such as elements of a stereotypical situation (e.g., going to a birthday party), as well as how to use the information, and what one can expect to happen next (e.g., in a restaurant we may expect to be seated and given a menu).

Finally, a KR is a **medium of human expression** or a language in which we say things about the world. The upshot of this characteristic is that we give consideration to questions such as: How well does the representation function as a medium of expression? How general is it? How precise is it? Does it provide for expression? These questions all relate to an overriding concern with the "ease of use."

Knowledge Representation Technologies

Several representations that have been designed with these characteristics in mind have emerged: **predicates**, **semantic networks**, **frames**, **cases**, and **scripts**. (Cognitive network topologies are discussed in Chapter 7, The Network Approach.) Basically, they differ in the ways that these representations are implemented (i.e., in their details). Each leads to a different set of properties that we may want a representation to have. Taken together, these comprise **KR technologies**. Each technology addresses the roles described above. KR technologies differ primarily in the emphasis that they place on the various objectives within KR.

Predicates

A **predicate** is an assertion of a fact about one or more entities or subjects. Predicates are just like the assertions that we make about subjects in grammar. Predicates, like all of the technologies currently in use, must be unambiguous, and word-sense ambiguity must be eliminated in order for them to be proper surrogates for the objects or ideas to be represented.

In the first instance, internal representations must eliminate all referential ambiguity. The computer must be able to correctly identify the "object" without confusion (e.g., one avoids such inputs as "He has a pain in his lower back"—who is the "he?"). This is not necessarily an easy thing to accomplish. For example, to represent the individual "Mary Johnson" in the machine, you could simply use "Mary," but there are many people with the name "Mary." You could choose "Mary Johnson," but once again there may be some number of "Mary Johnson(s)." An appropriate solution is to make up names. These

unique names are called **instances** or **tokens** (e.g., Mary-1, Mary-2, etc.). They are not English words but rather identification **symbols** and are useful for symbol manipulation, which is the computer's primary forte. These are surrogates for the person or object.

Suppose that the machine, after translating a natural language submission from a user, is "told" that:

"Mary-1 caught ball-10," and

"John-3 caught cold-5"

Subsequently, a user enters the query

"Who is ill?"

Because we have not further resolved the meaning of "caught," we might expect the machine to respond with

"Mary-1 caught ball-10 and John-3 caught cold-5"

To eliminate this word-sense ambiguity, we should introduce a distinct meaning for each instance of caught:

(a) catch-illness

(b) catch-object

The new internal representation (for Mary) would thus become:

"Mary-1 catch-object ball-10"

This appears to be a rather contorted representation of the simple fact that "Mary caught a ball," but without such formalities the machine becomes virtually useless as a symbolic processor.

Internal representations are called **formulas**, and when they are stored with other formulas, as in a database, are also called **assertions**. Thus, (Mary-1 catch-object ball-10) would be an assertion if it appeared with other formulas in a database.

Organization of the mind, particularly as it relates to how we represent the world, is a key concept in cognitive science. Recall from Chapter 1 that a basic form of mental representation included two fundamental elements: a **concept** or idea that represented a class of things that have been grouped together, and

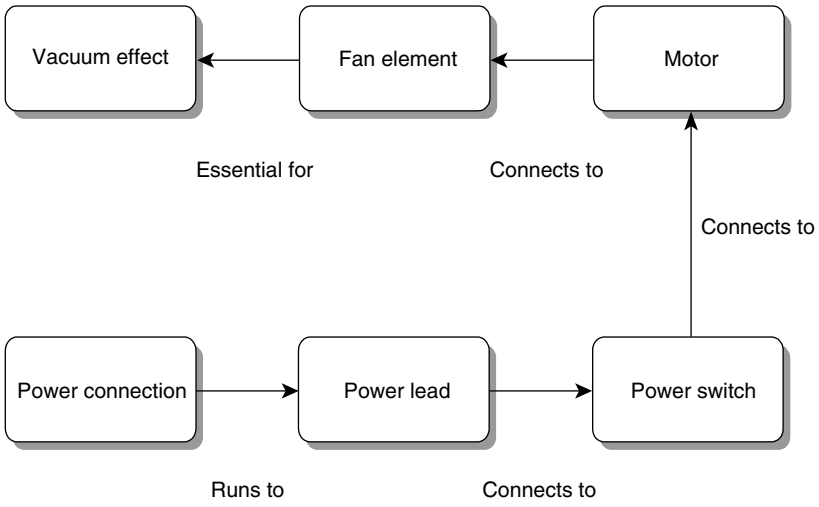


Figure 11.2 Simple semantic network for a vacuum cleaner

production rules or conditional statements (e.g., "If (x is true) then (y follows)," where "x" and "y" are propositions or predicates whose truth or falsity can therefore be determined). Concepts and rules are replaced by symbols that we use to reason and bear a resemblance to corresponding linguistic elements. Intelligence of this sort is reflected in language, and once we have a linguistic representation of an idea, reasoning about it becomes a process of symbol manipulation; consequently *machine computation can be considered to be a metaphor for reasoning*. A computer is an excellent machine for processing symbols. Computers may therefore be admirable candidates for representing intelligent agents (Poole et al., 1998). Formal representations of cognitive information (i.e., facts and rules) together with some form of logical reasoning constitute essential elements of an ES.

Semantic Networks

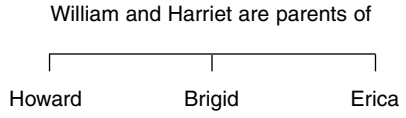
A semantic model of data representation includes **nodes** and **arcs**, where nodes represent particular concepts or elements of the world and arcs represent relationships between the concepts or elements. They are easily understood by humans and can be adopted for use in automated processing systems. A representative semantic network for a vacuum cleaner is shown in Figure 11.2.

With such an arrangement, an automated system could answer questions such as "How does power get to the motor?" and "What is the purpose of the

motor?" (Sowa, 1984). This network description has a number of similarities with the semantic networks as described in Chapter 7. That is, it includes nodes as well as arcs wherein the node's activity spreads outward to activate other nodes. In Chapter 7, semantic networks were used to describe informational organization within memory. (This discussion is a direct parallel with the material in that chapter, in which semantic networks include individual nodes that represent meaningful concepts and provide an understanding of the ways in which information can be retrieved from long-term memory.)

Frames

This is also a useful knowledge representation system with easy accessibility for humans and machines. Marvin Minsky—a name that appears repeatedly in the history of AI—introduced it. A **frame** can be considered a convenient way to represent a set of predicates. It will include a place to store attributes or properties of the concept (e.g., a **slot** for each attribute) and actions (i.e., **procedures**) that can be applied to the concept. As an example, suppose that we hope to represent a genealogical tree or semantic network as follows:



The **frame** describing William might be stored as:

William:

Sex: male

Mate: Harriet

Child: (Howard Brigid Erica)

In this case, sex, mate, and child are **slots**. To complete the genealogical tree, seven frames would be included: William, Harriet, Howard, Brigid, Erica, male, and female. One example of such frames is:

Sex (William, male)

Mate (William, Harriet)

Child (William, Howard)

Child (William, Brigid)

Child (William, Erica)

Each of the entries just noted is a predicate.

Cases

A **record** is an instance of an activity or event. Records are very similar to frames in that they require consistent index identifiers that allow cases to be compared to each other. A **case**—sometimes called a **record**—with information about a specific car sale would include data on the buyer, the specific product purchased, price, and so on. A car dealer could use a collection of such cases in many circumstances (e.g., identify market/product trends, etc.).

Scripts

Scripts seek to capture our expectations. They are similar to frames in that the object being defined is organized by attributes and associated procedures. The most important characteristic of a script is that it is applied to a particular situation or story. Thus, the ontological aspects of this technology are its important characteristics. The classic example of a script describes what happens when we go to a restaurant. We expect to see tables, chairs, utensils, waiters, and menus; we do not normally expect to see an ocean. When we attend a birthday party we do not expect to be eaten by a wild animal. A much-cited example follows along the following lines:

John went to a restaurant.

John ordered a hamburger.

When the hamburger came, it was overcooked.

John left the restaurant without paying.

A reasoning system would be able to infer that John did not eat the hamburger even though there weren't any details about what he did, or did not, eat (Schank & Abelson, 1977).

Frames as well as scripts provide a diverse and productive basis for representing clusters of knowledge about everyday situations and specialized situations. They "imitate" a recurrent feature of our own thinking processes; we

often interpret new situations on the basis of familiar stereotypes that provide us with the ability to build up an understanding of the circumstances without starting from scratch.

Machine Reasoning

Predicate Calculus

A sub-goal of "thinking machines" of the type being described (and of human cognition) is to draw inferences. To **draw an inference** is to come to believe a new fact on the basis of other information.

Predicate calculus is a way of calculating (by symbol manipulation) the truth or falsity of propositions and includes a language for expressing propositions and rules by which to infer new facts from those that are given.

The idea of a predicate calculus was first introduced in Chapter 1. Figure 11.3 is a diagram of the predicate language.

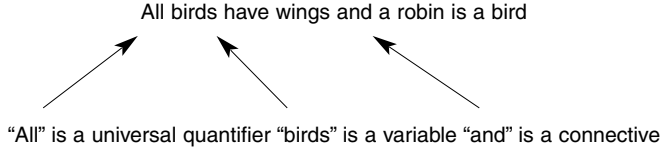
In addition to the elements shown in Figure 11.3, other resources permit us to increase the capabilities of predicate calculus. These include:

Connectives: We can build more complicated formulas by combining simple or "atomic" formulas using connectives (and, or, not, if [implication]).

A **variable** is an object or token.

Universal quantifiers certify that something is true for all possible values of a variable (e.g., "all (women)," or "all(x)" where "x" would be replaced by an instance of "women").

One way in which connectives are used is illustrated by the following example:



The predicate language is the basis for internal representations of the real world. As an example, suppose we want to represent a relatively simple set of circumstances; namely, that Professor Smith is currently located in room 300 (r300) of Memorial Hall. Figure 11.4 shows how this might be represented in three distinct contexts: conceptually, in Professor Smith's own mind;

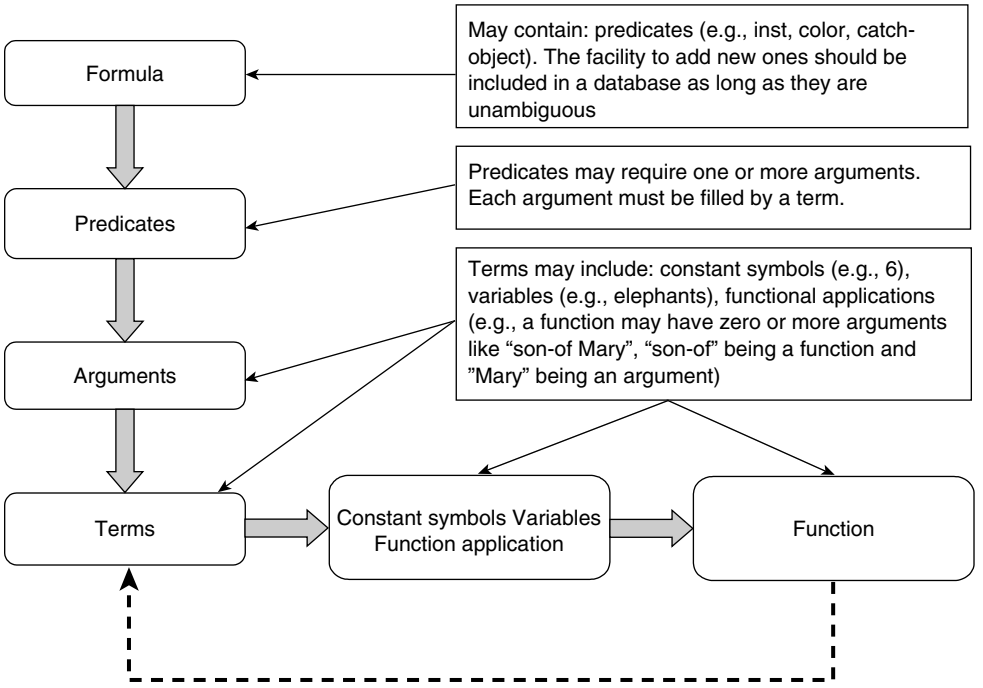

Formula
Predicates
Arguments
Terms
Constant symbols Variables
Function application
Function
May contain: predicates (e.g., inst, color, catch-object). The facility to add new ones should be included in a database as long as they are unambiguous
Predicates may require one or more arguments. Each argument must be filled by a term.
Terms may include: constant symbols (e.g., 6), variables (e.g., elephants), functional applications (e.g., a function may have zero or more arguments like "son-of Mary", "son-of" being a function and "Mary" being an argument)
Example: Turnip-1 tastes bitter.
Arguments
(Bitter (taste (turnip-1)))
Predicate
function
term (constant)

Figure 11.3 A Diagram of the Predicate Language

physically, in the real world; and within a computer that has been programmed to represent the same information.

There are three distinct representations in Figure 11.4:

- The physical view or facts and circumstances in the real world, including Memorial Hall and room r300, with the Professor in the room;
- The cognitive view or the Professor's self-perception of the situation represented in his mind, whereby he "understands" several abstract ideas

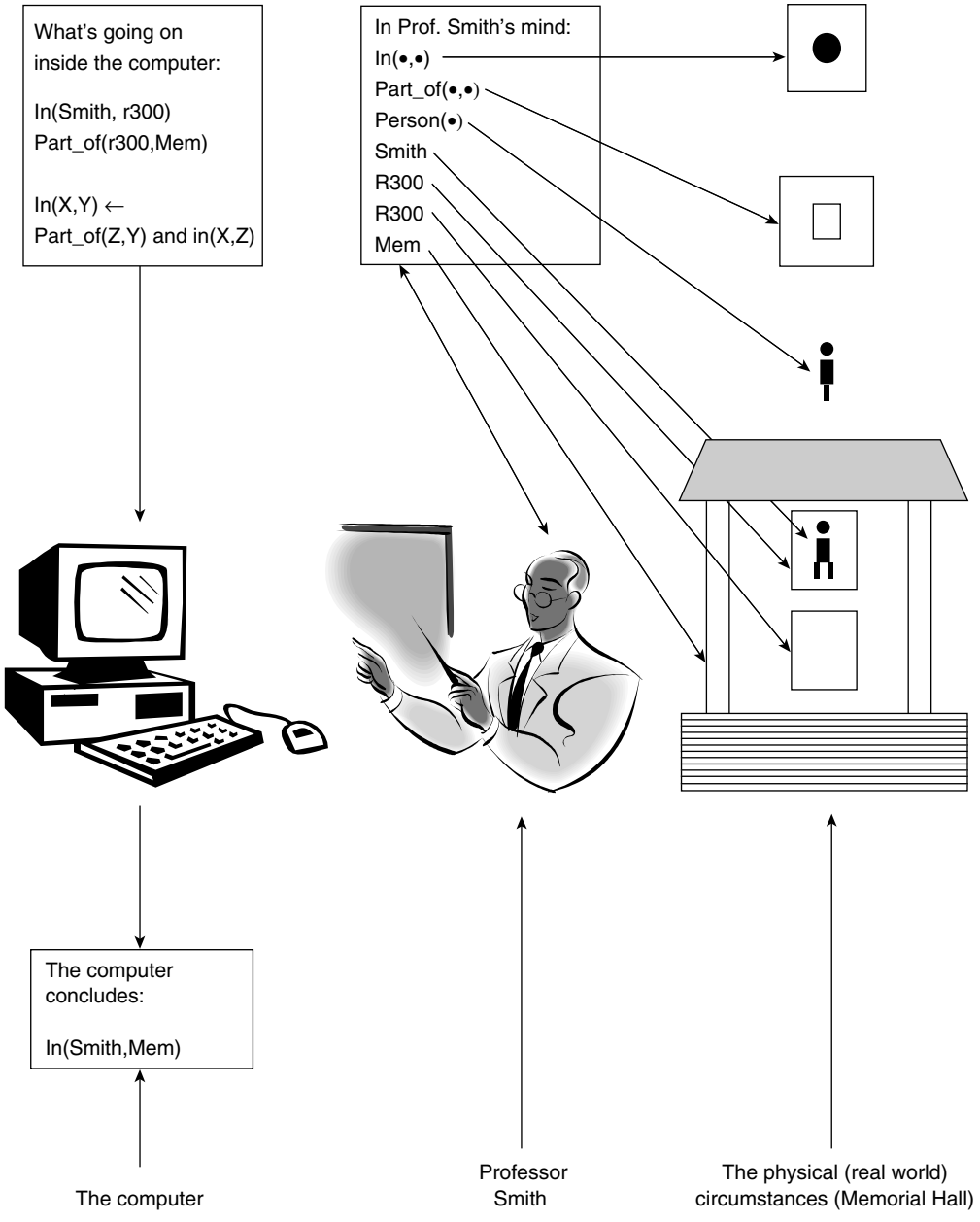


Figure 11.4 Relationship between physical circumstances, a mental model, and what is going on within a symbol-processing machine

representing the real world: the concept of a room; the concept of what it means to be within a room; the concept that a room can be part of or within a building; his self-awareness (i.e., consciousness of himself); his awareness of a particular room (r300) together with what makes this room unique among the general class of rooms; his awareness of the building (Memorial Hall).

- Finally, the computer representation of the facts that Professor Smith—a particular instance of something or someone being within another entity (r300)—is in room r300, and that room r300 is part of Memorial Hall ("Mem"). Once again this last fact is a particular instance of an entity "being part of" a second entity. In addition, the computer includes a rule; this rule states that

> if (an entity ("Z") is part of a second entity ("Y")
> **and** entity "X" is in entity ("Z")) then (one can
> conclude that entity ("X") is in entity ("Y"))

The computer does not have to be "told" that the Professor is in the room; it concludes that with its own "reasoning."

Logical Reasoning (Deduction, Abduction, Induction)

Logic is formal; truth or falsity of propositional statements can be inferred with given forms (**syntax** or linguistic elements) and logical manipulation. Using true or false outcomes, conjunction, and predication, an expressive logic can be built and more subtle ideas can be represented. A properly formed statement or **proposition** has one of two possible values, true or false. Simple examples of propositions include:

"Seven plus six equals twelve," and

"John is Mary's uncle."

Disjoint or stand-alone propositions are not very interesting. Take, for example, the proposition noted above, "John is Mary's uncle." While it may be true that John is Mary's uncle, we might have more information if we knew that "John drives a bus." This information could be represented as "John is Mary's uncle and John drives a bus." This enhanced information is obtained by using the **and** connective. Combining propositions with the connectives that are available to us provides the basis for predicate calculus. As shown in

Table 11.1 Truth Table for logical operators

Logical Value for X	*Logical Value for Y*	*And operation result*	*Or operation result*	*X implies Y result*	*Not X result*(1)	*X equivalent to Y*
T	T	T	T	T	F	T
T	F	F	T	F	F	F
F	T	F	T	T	T	F
F	F	F	F	T	T	T

(1) not operator includes only one operand (variable)

Figure 11.4, the computer can conclude that Professor Smith is in room r300; it does not have to be told this piece of information. Although this piece of information might have been entered into the computer as a fact, our machine would be quickly overwhelmed by the amount of data to be stored and would therefore not be particularly useful.

The connectives are critical to logical reasoning and are defined in Table 11.1. The table shows how two **variables**—objects or tokens, such as surrogates for Professor Smith—would be combined to compute a resultant logical value. To interpret entries in the table, consider the following example: If a variable having the identifier (name) "X" with a true ("T") value is combined with a variable whose identifier (name) is "Y" (also with a value of "T") using the **and** connective (operator), the resulting phrase will also be "T." Tables of this kind define the logical operators and are called **Truth Tables**.

An example that is closer to our everyday experience would be a proposition such as:

Clearance to enter the parking garage = (light is green) **and** (entry gate is up)

Following the entries in the Table 11.1, let X refer to the condition of the light. If the light is green, then X has a value of true (T); if the light is red, then it is not green and X (the condition of the light being green) is false (F). The same reasoning would be applied to the second variable—the condition of the gate. If the gate is up, the variable Y (representing the condition of the gate) is T; if the gate is down, then Y is F. The truth or falsity of the entire proposition "clearance to enter the parking garage" can be determined by following the entries in the table in the column marked "And operation result." The result of this analysis is repeated below, with X and Y replaced by their respective meanings.

Light green	gate up	Result of logical **and** operation (clearance to enter parking garage)
T	T	T
T	F	F
F	T	F
F	F	F

Thus, the only instance in which the logical result is true (based on the meaning of the logical **and** operation, as defined in the Truth Table) occurs when both the light is green (T) and the gate is up (T). All other conditions for the light and the gate result in a false state and "clearance to enter the garage" is denied (F).

The logical concepts depicted in the Truth Table may reflect the way in which our neurons are interconnected. If two neurons are arranged in serial fashion, then both must "fire" in order for the signal to pass to the next neuronal plateau. If the two neurons have a parallel arrangement, then the signal will pass to the next neuronal plateau if either one of the neurons "fires." These conditions are depicted in Figure 11.5.

Drawing Inferences

A goal of the designers of machine intelligence is to have those machines (of the type being considered here) **draw inferences**. To draw an inference is to come to believe a new fact on the basis of other information. There are several types of logical inference. We limit discussion to the most prominent of rule-based reasoning systems: forward reasoning or Forward Chaining, backward reasoning or Backward Chaining, and inductive reasoning.

Forward Chaining. This is built on deductive logic; it was originally cited in Chapter 1 and is repeated below for convenience.

If p then q

p is true

Therefore q is true.

A typical example of such logic from everyday experience is the following logical sequence:

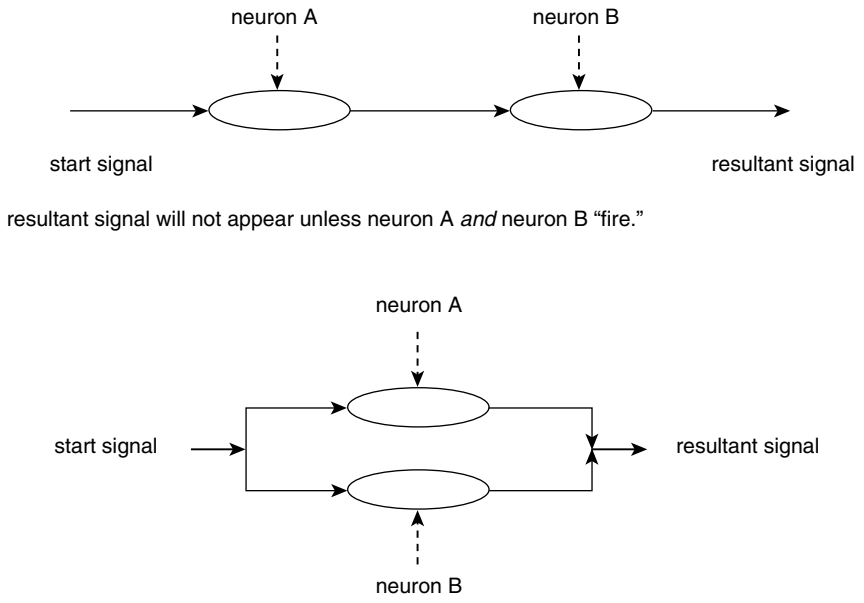

neuron A
neuron B
start signal
resultant signal
resultant signal will not appear unless neuron A *and* neuron B "fire."
neuron A
start signal
resultant signal
neuron B
resultant signal will not appear unless neuron A *or* neuron B (or both) "fire."

Figure 11.5 Neuronal organization for logical operations

All birds have wings

A robin is a bird

Robins have wings.

Our early development and education are built on such thinking and reflect the way we think in many situations—on the basis of facts that we have stored in our memory. Figure 11.6 is a schematic of the formal methodology that illustrates how Forward Chaining is accomplished.

Rule-based systems using Forward Chaining are "data-driven." Such systems follow the data antecedents (the "p" portion of the production rule "If p then q"), and determine the consequences or assertions that follow (the "q" portion of the production rule "If p then q"). The methodology used by Forward Chaining systems shown in Figure 11.6 is applied to a simple example, shown in Figure 11.7.

In this example we are given a number of facts, all pertaining to an automobile: the battery is good—fact A; the spark plugs are good—fact B; the car

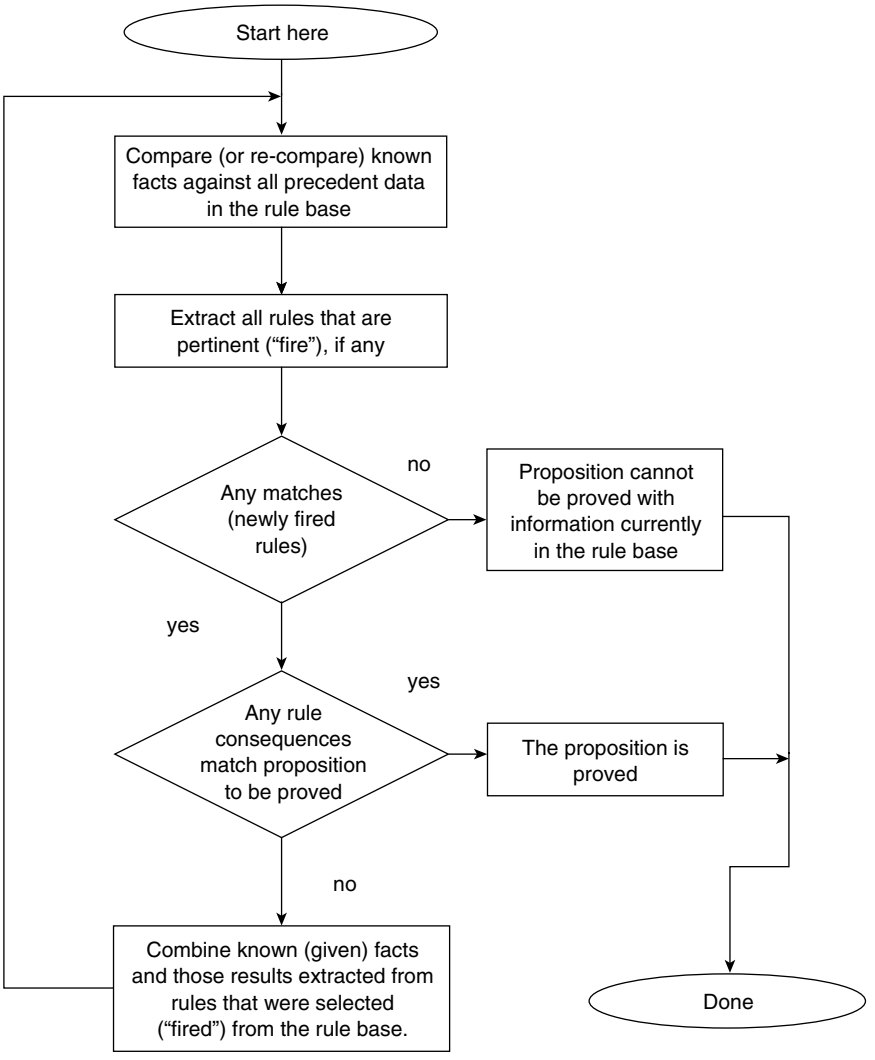

Start here
Compare (or re-compare) known facts against all precedent data in the rule base
Extract all rules that are pertinent ("fire"), if any
Any matches (newly fired rules)
no
Proposition cannot be proved with information currently in the rule base
yes
Any rule consequences match proposition to be proved
yes
The proposition is proved
no
Combine known (given) facts and those results extracted from rules that were selected ("fired") from the rule base.
Done

Figure 11.6 Simplified diagram of the methodology for Forward Reasoning (Chaining)

has gas—fact C; and the car has good tires—fact D. In addition, there are four production rules as follows:

Rule 1: If (the battery is good) then (the car has electricity).

Rule 2: If (the car has electricity **and** the spark plugs are good) then (the spark plugs will fire).

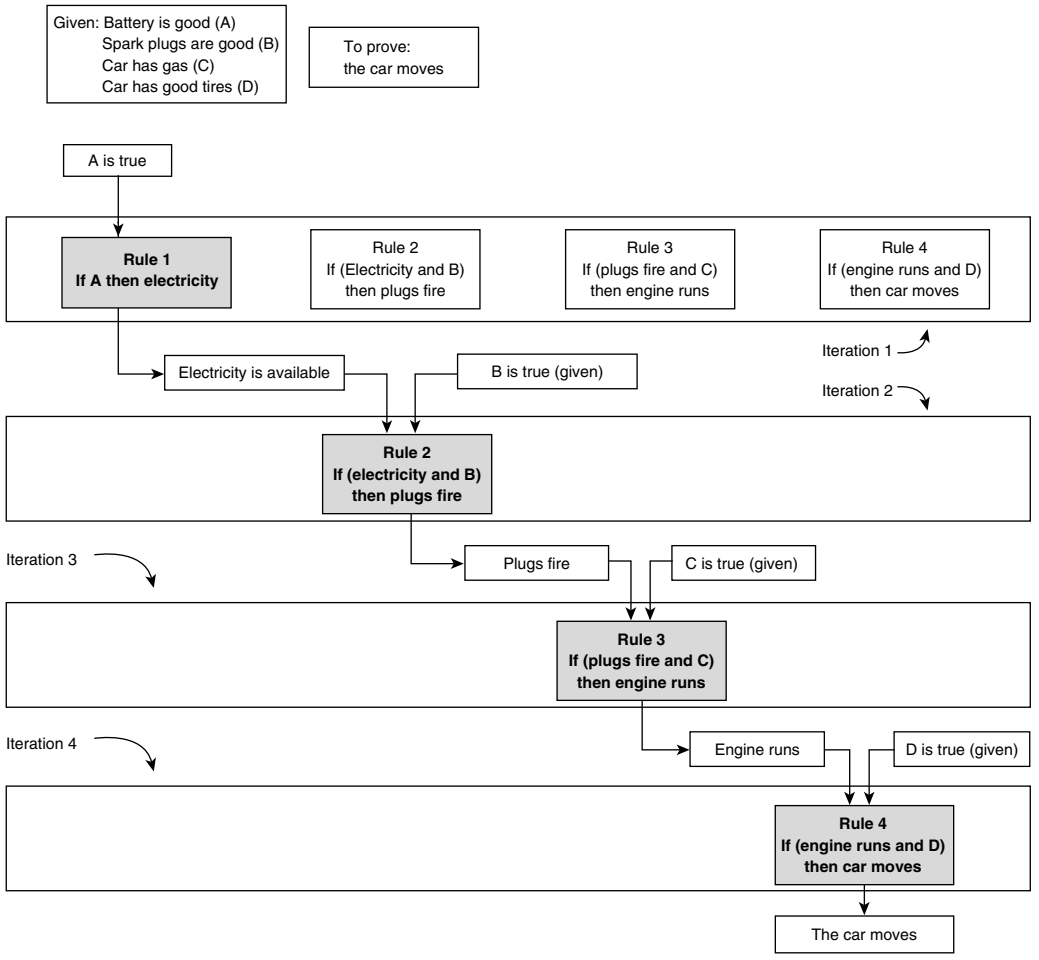


Figure 11.7 Example of Forward Reasoning (Chaining)

Rule 3: If (the spark plugs fire **and** the car has gas) then (the engine runs).

Rule 4: If (the engine runs **and** the car has good tires) then (the car moves).

It should be recognized that many other conditions must be satisfied for the car to move; this is a greatly simplified demonstration of Forward Chaining. A reasoning machine could have had one production rule that would have immediately confirmed the fact that the car would move. (The reader is encouraged to define this single rule.) However, by developing four production rules in

place of a single principle, a system takes on much greater versatility—as such rules can be used in other reasoning situations. Using forward chaining, a machine is able to prove that the car can move.

Forward chaining utilizes causative agents to establish their corresponding effects. Therefore, starting with fact A (that the battery is good), the machine would immediately explore Rule 1, because its causative phrase contains this fact. In other, more complex examples, fact A may appear in many production rules, and the machine would therefore have to examine each of these possibilities. In the present circumstances there is only one rule that has A as its causative phrase. The remaining sequence is readily followed in Figure 11.7.

Backward Chaining. Logical structures of this kind follow from **abductive** logic. The following sequence demonstrates abductive logic:

If p then q

q is true

Therefore p is true.

While it is easy to demonstrate that such logic is flawed, it is a kind of logic that we often use.

If (birds have wings) then (they can fly)

birds can fly

Therefore (birds have wings).

This type of reasoning is characteristic of abductive logic. Humans often use such reasoning when known facts do not immediately account for the observations. Therefore, we cannot apply forward (i.e., deductive) reasoning to draw conclusions. In such circumstances we seek to develop a theory for what we see. This is very characteristic of the way your physician develops a diagnosis. He or she evaluates the symptoms and then decides which disease or deficit is apt to be the source of these symptoms. Scientists use backward reasoning to develop theories that are consistent with the observations. Psychology is a science where this is a common occurrence.

Backward chaining starts with the desired goal. It then proceeds to find whatever evidence is needed to prove the goal. This evidence becomes what is called a **sub-goal**. (There may be more than one sub-goal.) These sub-goals lead to yet other sub-goals, until ultimate sub-goals are readily confirmed by facts.

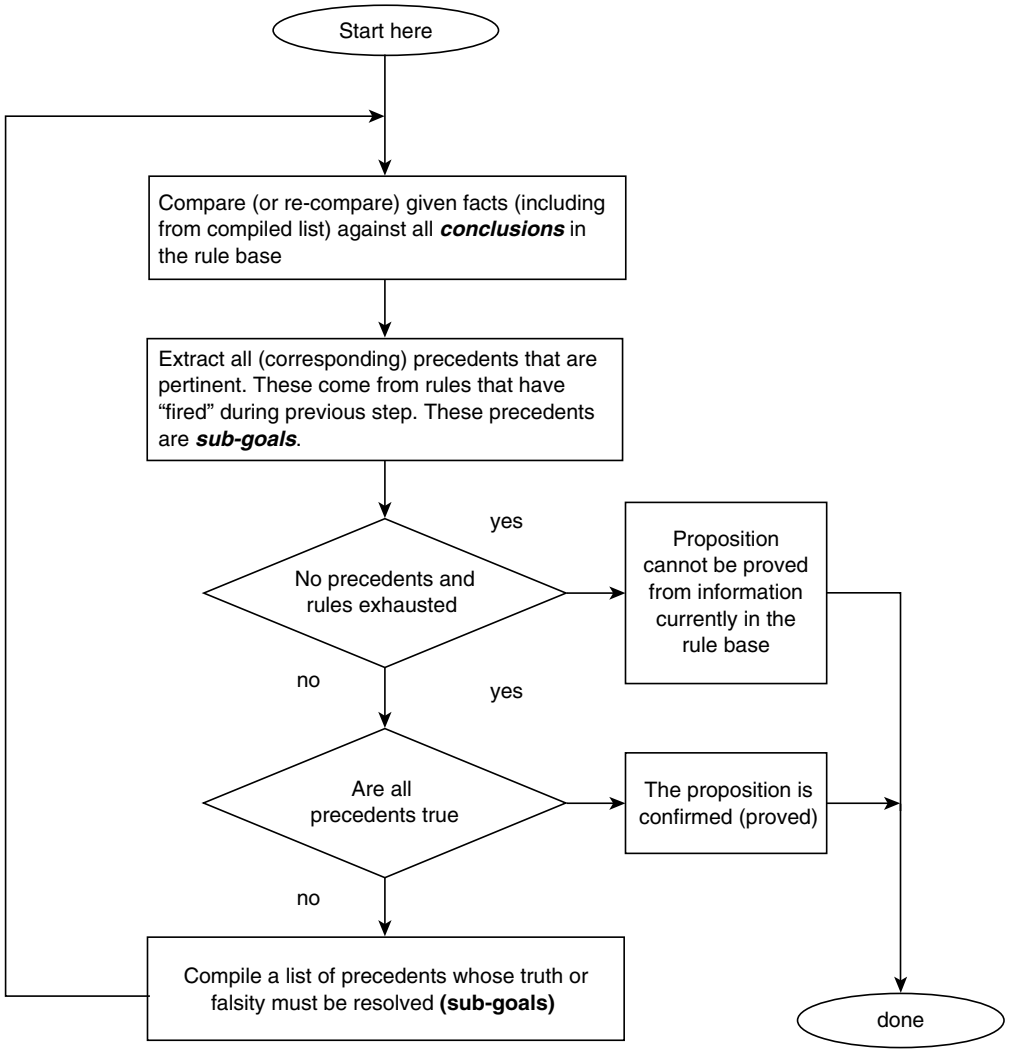

Start here
Compare (or re-compare) given facts (including from compiled list) against all ***conclusions*** in the rule base
Extract all (corresponding) precedents that are pertinent. These come from rules that have "fired" during previous step. These precedents are ***sub-goals***.
No precedents and rules exhausted
yes
Proposition cannot be proved from information currently in the rule base
no
Are all precedents true
yes
The proposition is confirmed (proved)
no
Compile a list of precedents whose truth or falsity must be resolved **(sub-goals)**
done

Figure 11.8 Simplified diagram of methodology for Backward Reasoning (Chaining)

By "reversing the chain" of circumstances that have been resolved, the original proposition is ratified. A simplified diagram of the methodology for such reasoning is shown in Figure 11.8, and a solution corresponding to the problem diagrammed in Figure 11.7 is found in Figure 11.9. Backward Chaining

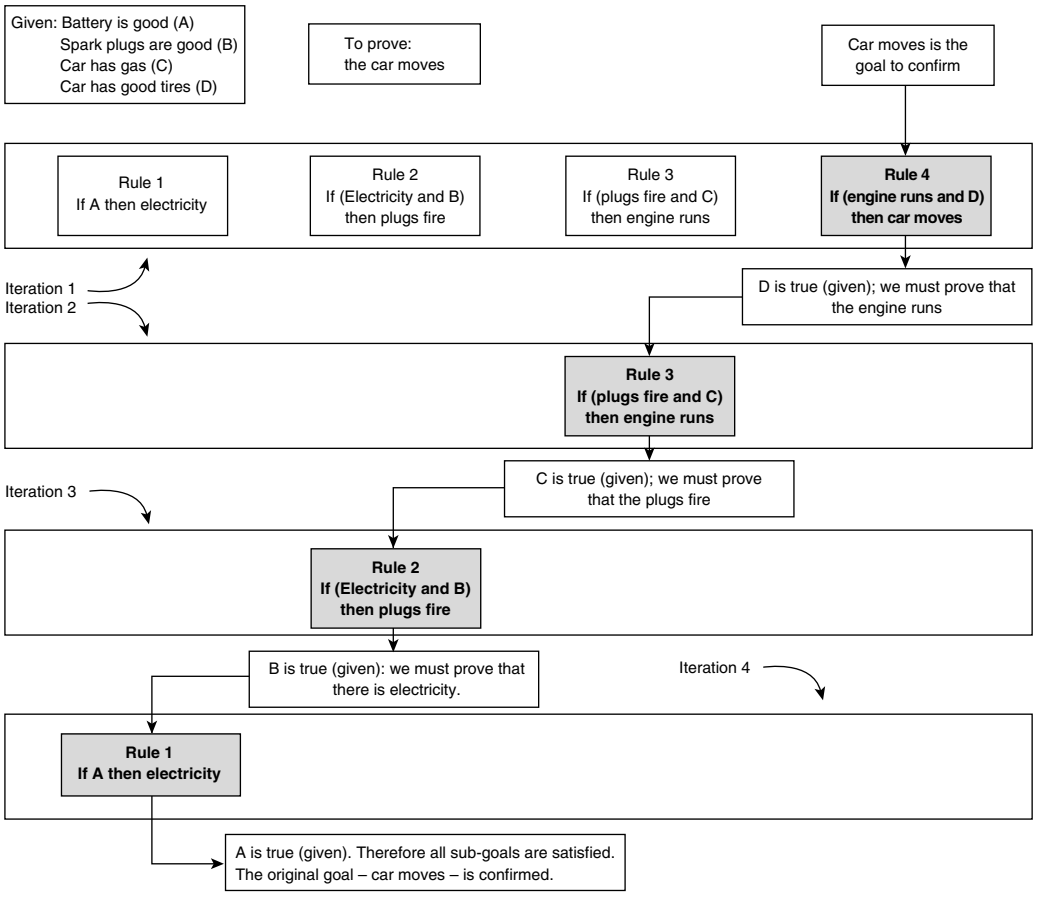

Given: Battery is good (A)
Spark plugs are good (B)
Car has gas (C)
Car has good tires (D)
To prove:
the car moves
Car moves is the goal to confirm
Rule 1
If A then electricity
Rule 2
If (Electricity and B) then plugs fire
Rule 3
If (plugs fire and C) then engine runs
Rule 4
If (engine runs and D) then car moves
Iteration 1
Iteration 2
D is true (given); we must prove that the engine runs
Rule 3
If (plugs fire and C) then engine runs
C is true (given); we must prove that the plugs fire
Iteration 3
Rule 2
If (Electricity and B) then plugs fire
B is true (given): we must prove that there is electricity.
Iteration 4
Rule 1
If A then electricity
A is true (given). Therefore all sub-goals are satisfied.
The original goal – car moves – is confirmed.

Figure 11.9 Example of Backward Reasoning (Chaining)

starts with the effective phrases of production rules and establishes the corresponding causative phrases of those rules as one or more sub-goals. The machine continues this process until all sub-goals have been confirmed, thereby establishing the original target goal.

It is to be noted that in Figure 11.9, Rule 4 contains the goal (i.e., the car moves) that we seek to establish. The machine thus proceeds to establish that the "engine is running" and that "the car has good tires." This latter element has been posted as a fact; therefore the machine must determine that "the engine is running." As the reader can readily determine, this can be accomplished using Rule 3 and its consequences.

Inductive Reasoning

One final reasoning mechanism worth noting is that of induction. The most common form of this is:

From: (P a), (P b), (P c), . . .

Infer: (for all (x) (P x))

(P a), (P b), (P c), and so on all signify that objects whose properties are a, b, c, and so on belong to the category identified as P. We therefore conclude that any object whose properties are similar to those of a, b, c, and so on belong to the category identified as P.

All of this is best illustrated with a simple example.

From: (if (inst leaf-1 leaf)(color leaf-1 green))

(if (inst leaf-2 leaf)(color leaf_2 green))

Infer: (forall (x) (if (inst x leaf)(color x green)))

Using many observable instances that have to do with leaves, the machine reasoning system has concluded that leaves are green. This type of reasoning is known more commonly as "learning." Learning is a fundamental competence among our cognitive abilities. It has proven to be one of the most difficult things for computers to do.

Evaluation of Rule-Based Reasoning

Abstracting the real world has a number of inherent difficulties. For reasoning and communication to be robust, we must be able to translate or map facts into equivalent internal machine representations and back again so that a user can readily interpret the information. In a system, we might map "Spot is a dog" into (inst spot-1 dog). If, somewhere in the reasoning system, we had the additional fact that all dogs have tails, then, using deductive logic, we could conclude that Spot has a tail. Internally this might appear as ("hastail spot"). If the mapping function were reliable, a user would see "Spot has a tail." However, converting natural language facts into internal representation may not be so easy. Consider two statements:

1. All dogs have tails.
2. Every dog has a tail.

Statements 1 and 2 might represent the same information to humans. However, there might be two internal interpretations for statement 1. It might mean that every dog has at least one tail; it might also be internalized to mean that each dog has several tails.

Procedures for resolving axioms (e.g., theorems) may not come to a computational conclusion if the proposed statement is not a theorem. Reasoning systems must make provisions for avoiding such difficulties. These may include the imposition of a limit to the number of iterations that can take place in the absence of finding an answer or a limit to the amount of time allotted to the query.

When machine intelligence is built on simple facts as well as the inference mechanisms described above, useful results are possible. Unfortunately, the real world is not easily described using these limited representations. For example, they do not indicate how one object is related to other objects. In addition, a description of the world needs to include not only "static" descriptions but also "dynamic" representations or typical sequences of events.

Expert Systems

KR and machine-based reasoning methods described above constitute the foundations for Expert Systems (ESs). An ES uses information that is restricted to a specific knowledge domain and is dedicated to solving problems and giving "expert quality" advice to users in that application area. It originates from the "top-down" paradigm of AI; it is the oldest and most advanced area of AI and has a successful record of commercial application, as well as a record of achievement that frequently matches or exceeds the performances of human experts. Figure 11.10 depicts the modules and organization of an ES.

The designer of the ES will accumulate information for the KB through discussions with human experts. The designer, often referred to as the Knowledge Engineer, must be skilled in information technology, as he or she must limit or filter the intelligence until it becomes intelligence that represents the essential features of the specific domain. As the system matures, additions and subtractions to the KB are standard practice. The KB is a complement of both the theoretical and practical (heuristic) information. Information of the latter variety includes: "tricks," shortcuts, and so-called "rules-of-thumb" accumulated by experts via problem solving experiences. An ES will emulate the methodologies of its human counterpart.

Expert Systems in Action

Do you think that it is possible for a computer to help you determine what's wrong with your vacuum cleaner? You may be familiar with call-in centers for

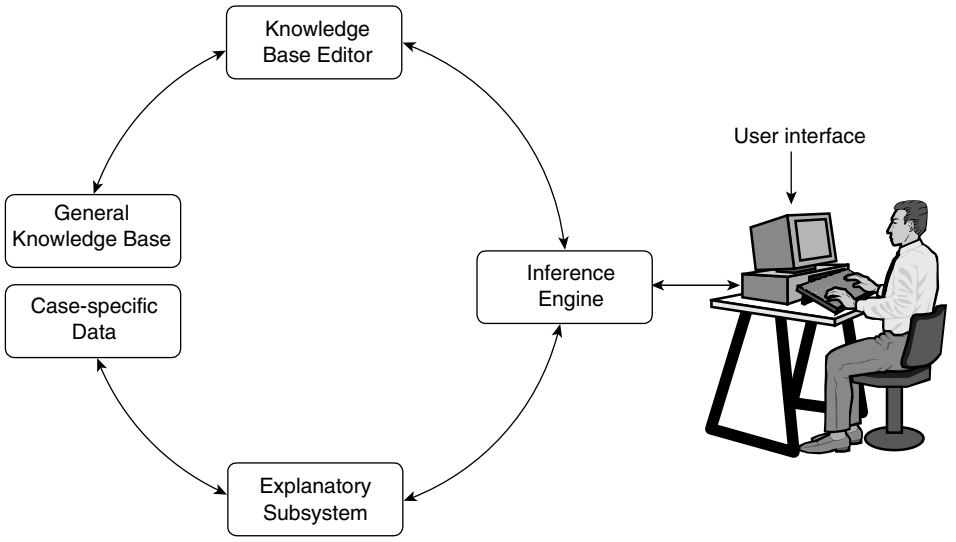


Figure 11.10 Organization of an Expert System

people with technical problems seeking advice. Figure 11.11 depicts a very simple KB for analyzing problems associated with vacuum cleaner problems.

The KB has a tree-like appearance and node A is considered to be the **root** of the search **tree;** all searches start at the root node. (By virtue of tradition, the root is at the top rather than at the bottom of the tree.) Terminating nodes (e.g., N, O, L, and M) are considered to be the **leaves** of the tree. An ES can examine this structure in one of two ways.

Breadth-First Searching. In this mode, the searching for proper advice to give the user starts with the root node and descends level by level. All nodes at a particular level are tested before the search proceeds to the next level. In this example, breadth-first searching would examine nodes A, B, C, D, E, F, G, H, I, J, K, L, M, N, and O, in that order.

Depth-First Searching. An alternate search scheme goes deeper into the search space whenever it is possible. Only when there are no dependents (leaves) of a state to be evaluated are sibling nodes—those on the same level—considered. In the current example, searching would test the nodes A, B, E, G, I, K, N, O, J, L, M, H, C, and finally F. Depth-first search systems use backward reasoning; the process is called **back-tracking**. When a "NO" node is reached and there are no further ways to further descend the tree, the search will proceed to the next accessible branch of the tree.

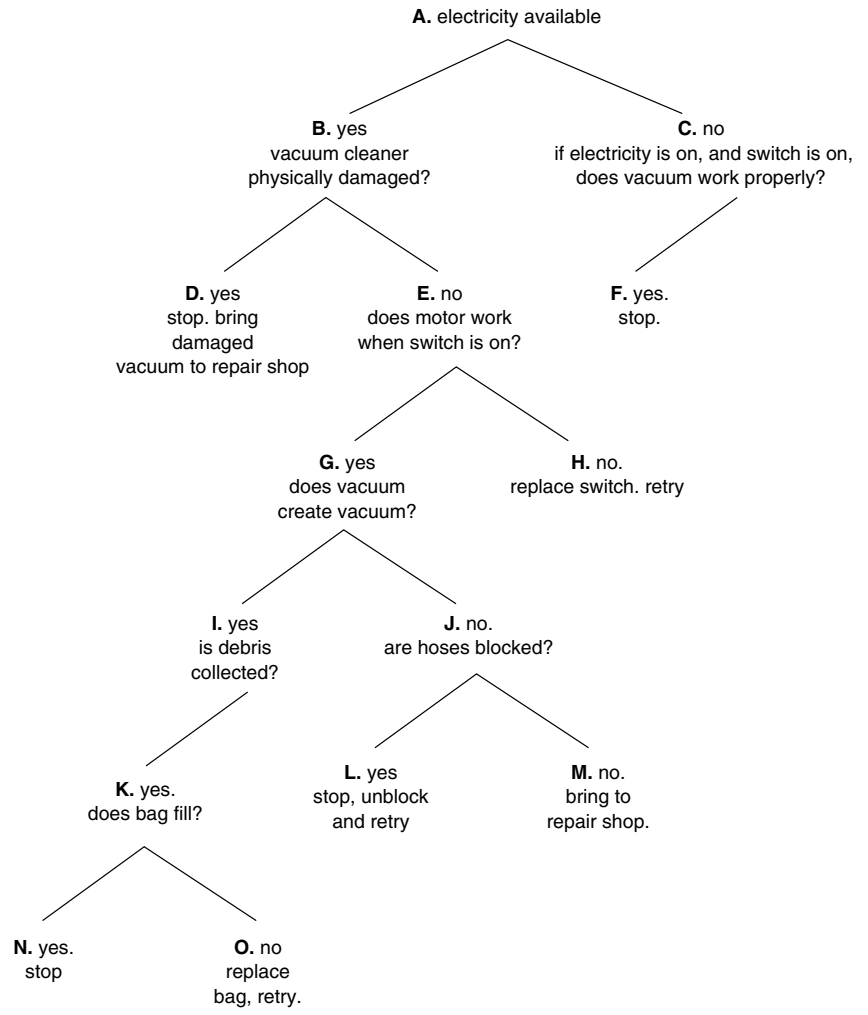


Figure 11.11 Diagram of a simple Expert System for analyzing vacuum cleaner problems

Breadth-First versus Depth-First tradeoffs. There are no simple answers to the Breadth-First versus Depth-First questions. The Breadth-First search scheme seeks to explore all analyses in parallel until they either succeed or fail. The machine, using a Breadth-First scheme, must keep a number of alternatives "in mind" concurrently and will require more memory space in the computer, and thus Breadth-First is more difficult to implement from a programming perspective than the Depth-First alternative. However, Breadth-First parsing generally gets to the shortest solution more efficiently than Depth-First programs. Depth-First searches are easier to implement. However, a Depth-First search

can become time-consuming when it is faced with problems of ambiguity. Search times are prolonged because constituents are built that do not contribute to the final solution. Partial solutions need to be undone and possibly rebuilt several times as backtracking develops. Broadly speaking, if we include the fact that one is interested in general knowledge, then a Breadth-First approach is signaled, whereas interest in specific knowledge might lead one to a Depth-First examination of the KB.

MYCIN

A far more interesting ES is among the first to have been developed—namely, the MYCIN system. It was first developed at Stanford in the 1970s and it has had a significant impact on the design of commercial ESs. Physicians have been reluctant to use MYCIN because of the ethical and legal implications of computers in medicine. (If something goes wrong, "whom do you sue?") On some occasions, however, it has outperformed faculty members of the Stanford Medical School. MYCIN modeled the (infectious) world as a set of IF-THEN rules with accompanying certainty factors. One of the rules, formulated in natural language, went as follows:

IF the infection is primary bacteremia

AND the site of the culture is one of the sterile sites

AND the suspected portal of entry is the gastrointestinal tract

THEN there is suggestive evidence (certainty 0.7) that the infection is bacteroid.

MYCIN was developed using the LISP programming language and is basically a goal-directed system that relies on the backward reasoning strategy. A series of heuristics that was superimposed on this structure was used to manage the search for a solution to the inquiry. An interactive fragment from a contemporary medical ES that is similar to MYCIN is shown below.

(Doctor)

```
Please enter findings          ;The program asks for facts about the
*sex male                      ;patient.
*race white                    ;There is a fixed vocabulary of symptoms
*alcoholism chronic            ;that must be followed
*go                            ;This instructs the ES to proceed
```

```
Disregarding:                ;The ES finds a set of suspected diseases
Exposure to rabbits          ;Symptoms not explained by these diseases
Leg weakness                 ;are put aside.
Creatinine blood increased
Considering:                 ;The ES explains its reasoning
Age 26 to 55
Ruleout:                     ;and rules out certain disease.
Hepatitis chronic
Alcoholic hepatitis
Abdomen pain generalized?    ;It requests additional information to
*no                          ;further refine its findings.
Abdomen pain right quadrant?
```

(*indicate physician responses, what follows ";" are explanatory comments)

This fragment does not represent a complete interactive session. The user would need to supply additional information to obtain a potential diagnosis. An excellent demonstration of such systems can be found on the Internet:

http://dxplain.mgh.harvard.edu/dxp/dxp.sdemo.pl?/login=dems/cshome

Evaluations of Expert Systems

Although ES technology has led to many useful applications, it was early on confronted with an inherent problem—how to determine what is "true" and what is "false." Contemporary ESs take this into account in a variety of ways; it remains one obstacle that greatly complicates application software. An ES is built on reasoning that relies on a human being or a machine's being able to conclude with certainty whether a phrase is true or false. An example: How do we evaluate the term "too fast" in an expression such as "the train is approaching the station at a rate that is too fast"? ESs as well as humans require a definitive answer, even though there may not be one. There are ways to deal with this within an ES, for example, using a probabilistic approach or a heuristic (i.e., judgmental) alternative.

ESs are limited by the information in their KBs and the process for putting that information into the KB. ESs cannot report conclusions that are not already implicit in their KBs. The trial-and-error method of eliciting information, programming, and testing is likely to produce inconsistent and incomplete KBs. An ES may exhibit important gaps in knowledge at unexpected times. ESs are unlikely to have complete, clear functional specifications, and designers are unable to predict their behavior in situations that are not fully

tested. Testing of KBs of any useful size is realistically not possible. Large KBs may be difficult to modify and maintain.

Software that is heuristically developed (i.e., via trial-and-error) is less reliable than software developed from precise specifications. To produce a reliable program that solves a problem requires one to study the problem. This is not an easy task because we cannot rely on people who simply "disclose the way that they solve it." As an example, consider that to distinguish among objects in a picture, one studies the characteristics of the objects as well as aspects of the photographic process (e.g., lighting, orientation, etc.). If the designer simply asks "experts" for the rules, a reliable program is an unlikely result. The expert may not be truly aware of the methods that he or she uses. The knowledge engineer is like an investigative reporter—and from your own experience you know that such people do not always ask the "right questions." Testing is an unreliable way to find errors in programs. Such testing may reveal the *presence* of problems or **bugs** but cannot reveal the *absence* of bugs. Furthermore, the "correct" behavior of an ES in a particular situation is often a matter of opinion. While ESs cannot be trusted as autonomous agents, they can be particularly useful as "intelligent assistants," especially when accompanied by human supervision. We should not ascribe "thinking" to ESs as they do not possess the same sort of "intelligence" as that of a human being. They can, however, "imitate" a number of processes that humans employ when solving problems. Peter Jackson (Jackson, 1999) has explored this in greater detail.

Counterarguments. Complete testing of an ES may not be feasible. The best that we can hope to achieve is a system that imitates human experts; such individuals are themselves imperfect. One does not expect a physician to correctly diagnose every set of pathologic symptoms. Mathematical proofs for large programs are meaningless because one cannot tell whether the specifications are complete or accurate. A prudent alternative with useful applicability: We can, and often do, build machines that overcome human limitations.

One of the great advantages possessed by the modern computer is its ability to explore many alternatives in a short time. Based on its evaluations of these alternatives, the machine can prioritize choices similar to those that experts are likely to suggest. One may conclude that whatever "artificial" mechanism such programs employ, the results support the idea that they may be imitating "thinking" processes within the human expert.

The concept of precise specification for human processes may itself imply an oxymoron—what is it that an expert does anyway? Despite the presence of numerous limitations, a number of highly successful ESs have evolved. The most successful are ESs based on well-understood models of diagnostic

processes that are used for diagnosing faults. The rules can be organized into a hierarchy of modules that correspond to decision points in the diagnostic process, as in MYCIN. Their completeness and consistency can be evaluated with respect to the model. The goal is to devise models of other problem-solving processes and to use them in a similar (hierarchical) fashion. AI cannot yet replace real intelligence. Alan Perlis of Yale notes, "Good work in AI concerns the automation of things we know how to do, not the automation of things we would like to know how to do."

Fuzzy Logic

Fuzzy Logic has emerged as a way to deal with some of the criticisms that have been leveled against ESs. Natural language is the embodiment of cognition and human intelligence. It is very evident that natural language includes an abundance of vague and indefinite phrases and statements that correspond to imprecision in the underlying cognitive concepts. Terms such as "tall," "short," "hot," and "well" are extremely difficult to translate into knowledge representation, as required for the reasoning systems under discussion. Without such precision, symbolic manipulation within the computer is bleak, to say the least. However, without the richness of meaning inherent in such phrases, human communication would be severely limited, and it is therefore incumbent upon us to (attempt to) include such facility within reasoning systems such as ESs.

In 1965, Lofti Zadeh (Zadeh, 1965), building on historical developments in mathematical logic, developed a concept he christened **Fuzzy Logic** (Lukasiewicz, 1967; Black, 1937). It gave rise to a new category of AI automaton— the Fuzzy Logic machine. The defining term represents an obvious extension of machines built on formal models of logic; decision making, particularly for humans, is "fuzzy" in nature. Like ESs, it is related to the "top-down" model but provides for greater flexibility in defining elements in the worldview. Fuzzy Logic provides us with a facility that enables a relatively precise measure of such information. The methodology permits us to translate measurements from the real or physical world into equivalent linguistic representation, to solve the resulting "word" problems, and finally to develop a precise or "crisp" result on which we can act. The Fuzzy Logic process is depicted in Figure 11.12.

Fuzzy Logic systems, as depicted in Figure 11.12, have value for both continuous-control and discrete-input applications. Continuous situations require the machine to apply the methodology in an ongoing "continuous" manner. Examples of continuous control applications include: managing the level of chlorine in a public swimming pool, maintaining the stability of an ocean liner, providing "zero-gravity" control of a passenger train so that it starts or stops

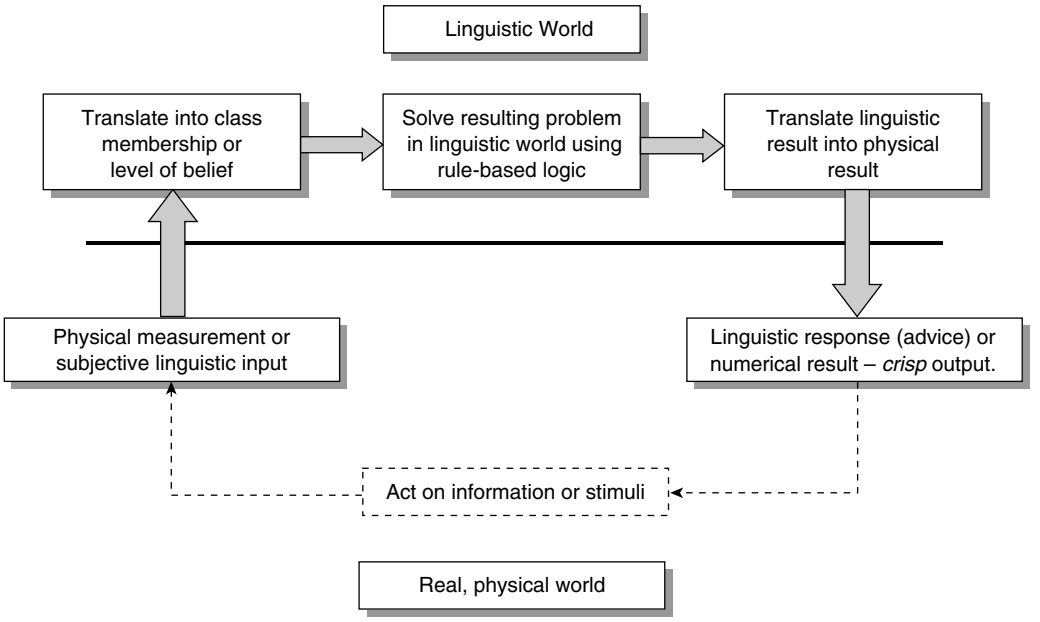

Linguistic World
Translate into class membership or level of belief
Solve resulting problem in linguistic world using rule-based logic
Translate linguistic result into physical result
Physical measurement or subjective linguistic input
Linguistic response (advice) or numerical result – *crisp* output.
Act on information or stimuli
Real, physical world

Figure 11.12 The Fuzzy Logic Methodology

without jolts, and maintaining the level of anesthesia during a surgical operation. These are all challenging tasks for a traditional engineering control system because they do not respond well to sudden changes in the physical world; nor can they take into account subtle distinctions such as the one embodied in the phrase "too fast" (as in the case where we wish to control the starting or stopping of a train). With respect to discrete-input examples, Fuzzy Logic can be an integral part of an ES in that it can "interpret" the expert's view of the world. For example, as part of proper medical care, the strength of a prescription will bear a strong relation to the physician's sense of the relative "strength" of the patient's symptoms, as well as the physical characteristics of the patient (e.g., weight, age). Fuzzy Logic can greatly enhance the richness of the data as well as the manipulation of the information when a physician queries the Fuzzy Logic Expert System.

Representation of Information in the Fuzzy World

When representing subjective or relative information in a Fuzzy computer, we assign a numerical value to the "**degree of belief**" that we believe describes the level of significance or validity of a specific property of an object that is

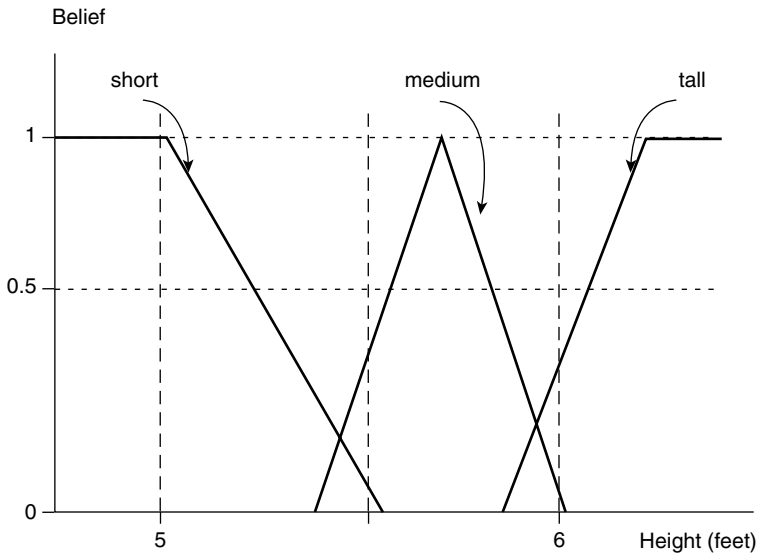


Figure 11.13 Membership classes for a tall, medium, or short person

part of the KB. Each datum represents a single instance of the broad class of knowledge to which it belongs. Thus, when referring to a person's height, we might use the terms "tall," "medium," or "short" as qualifiers or classifications of a person's height. However, linguistic richness is increased if we can further qualify each of these descriptors, for example, "not too tall." All such descriptions of "being tall" can be transcribed into a series of "membership classes." Although we can define a great many such classes for the sake of greater resolution and linguistic refinement, Fuzzy Logic systems may exhibit optimal performance when these classes are restricted to some parametric maximum. Some Fuzzy application software suggests a maximum of five such classes for each object being defined. Figure 11.13 shows membership functions reflecting a belief system in which a person may be tall, medium, or short, according to his or her height.

Consider the figure's competing interpretations of shortness. Each choice represented in the figure is simply one of many possibilities. Each of us may have a different interpretation of what it means to be short, medium, or tall. One of the challenges that confront Fuzzy systems is the need to determine class membership, and only slightly different membership functions will produce significant differences in a system's responses. There are several alternatives for determining class membership functions (e.g., neural networks can be

Table 11.2 Fuzzy Logic Rules

Membership Function A Belief	*Membership Function B Belief*	*Connective*	*Result*
x	y	and	Minimum of x or y
x	y	or	Maximum of x or y

used), but this level of technical complexity will not be introduced in this text. From the figure we note a very strong or "certain" belief—given a belief or membership value of 1.0—that people below five feet are "short." As we go from a height of five feet to a height of six feet, we encounter decreasing belief that a person is short. When we go above five-and-a-half feet, we totally disbelieve—a belief or membership value of 0.0—that a person is considered to be "short." Corresponding interpretations are applied to "medium" and "tall" people.

Fuzzy Logic Rules

Having described how nonprecise information can be represented in a machine in a quantitative manner, we introduce the Fuzzy Logic rules that underlie rule-based decision making in such automata. These are modifications of the connectives, as defined in the Truth Table (see Table 11.1, above), and represent the contribution that Zadeh made to AI in 1965. Despite the fact that such thinking has been around for a long time, it is a relatively new field of AI, and the influence of Fuzzy Logic in our world is just beginning to find voice as well as application. Table 11.2 summarizes the connective rules of Fuzzy Logic.

Notice that the traditional **and** connective operation no longer produces a true (1) or false (0) result. Instead, when connecting two Fuzzy variables using the **and** operation, the result will be the *smaller of the defining belief values*. The result for the **or** connective is the *larger of the defining values*. For example, in determining if a six foot person is "medium **or** tall," we get a result of 0.3, because our belief that a person who is six feet tall is "medium" has a value of 0.0, whereas our belief that a six foot person is "tall" has a value of 0.3 (approximately), as determined from Figure 11.13. The larger of these two values is 0.3, as required by the Fuzzy **or** operation. The Fuzzy rules as shown in Table 11.2 follow from mathematical concepts that are beyond the limits of our present discussion. What is significant from the AI perspective is the application of these rules, as shown below.

Fuzzy Advice for Decision Making or Management

A straightforward example will provide a small taste of the potential power of Fuzzy Logic (Klir & Folger, 1988). As may happen to you at various points in your life, you could be deciding which of four job offers to accept. Your decision may rest on several factors: salary is a strong consideration but other things may also matter. These include your interest in the job and the job's location (travel convenience). You have assigned ratings for these two characteristics to each of the four opportunities and generated a class membership profile. These are summarized in Figure 11.14. In this case the membership functions do not have a continuous appearance; instead, each of the parameters has a discrete value for the system of beliefs. For example, the salary membership value—or merit ("attractiveness")—for job offer number 1 has a value of 0.875, and the interest in job offer 3 has a value of 0.8. Evaluating the following function would tell you which offer to accept:

Accept an offer = (job attractiveness **and** job interest **and** job convenience)

For each job offer we choose the smaller membership value of the two, as that is what the Fuzzy rule for the **and** operation specifies. These results are sketched in Figure 11.15.

Figure 11.15 reflects the "verbal" answer to the decision-making problem. That is, the answer reflects a measure of the extent to which each job offer belongs to the class of "acceptance of the offer." We must now specify a way to convert this verbal offer (membership) into an action. There are several techniques for doing so in Fuzzy Logic systems. One of these methods—and the one that we will discuss here—is choosing the "most likely value." In this case the system recommends choosing job offer 2. Notice that this is not the offer with the highest salary. The other factors had an influence on the ultimate decision.

Evaluation of Fuzzy Logic Systems

Not everyone has embraced Fuzzy Logic; many critiques have arisen in response to the exuberant suggestions from its defenders about its potential. Historically, one of its severest critics has been Susan Haack (Haack, 1979). Haack argues that there are very few true decision-requiring candidates for which Fuzzy Logic is a potentially useful resource. The computer complications that follow from Fuzzy reasoning calculations therefore serve no useful purpose. (The job offer example provided above could have been resolved using principles from probability.) Haack argues that Fuzzy Logic depends on "truth" and "falsity" being continuous values and not discrete ones; if

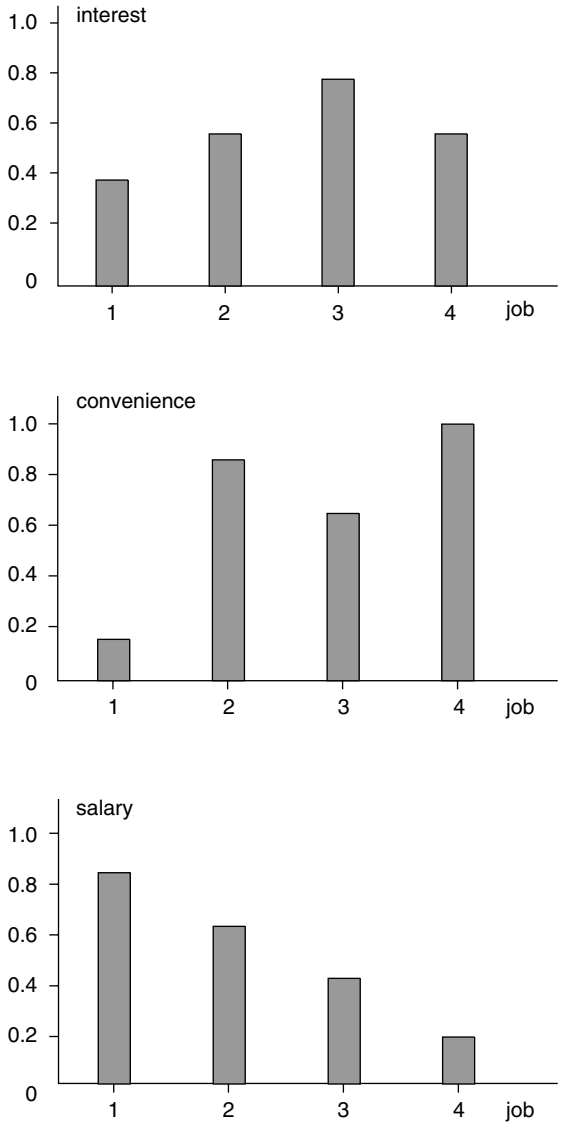


Figure 11.14 Membership classes for the Fuzzy Logic employment example

continuity were the case, Fuzzy Logic might be justified. However, Haack maintains that truth and falsity are indeed discrete concepts. For example, a statement such as "the sky is blue" is either true or false. In Fuzzy Logic, as well as linguistically speaking, one would refine the concept of a "blue sky"

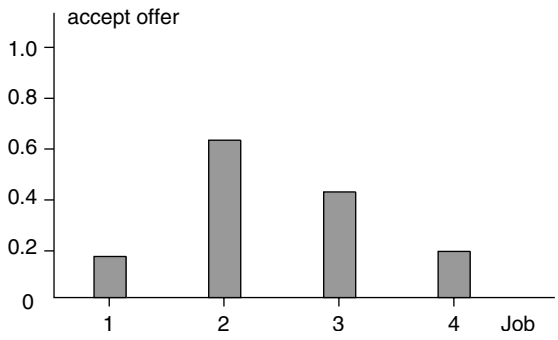


Figure 11.15 Fuzzy evaluation of the employment example

(e.g., "how blue"). For Haack, a "blue sky" is simply a matter of definition and not "degree."

J. Fox (Fox, 1981) has rebutted this line of reasoning by noting that there are three broad usages wherein Fuzzy Logic can produce benefits: in describing real-world relationships that are "fuzzy," as a prescription for those cases in which data are fuzzy and therefore require fuzzy calculations, and in instances in which some inferential circumstances are inherently fuzzy. Fox also notes that Fuzzy Logic and classic logic need not be competitive. Rather they can be complementary. Many of Haack's objections originate from deficiencies in semantic clarity. Ultimately, Fuzzy statements may be translatable into phrases that classical logicians would accept.

Fuzzy Logic supports so many practical applications with great success that it is likely that machine development in this area will continue at an accelerated pace. Fuzzy Logic has been shown to be potentially useful in ESs, in part because the domains of such systems are inherently fuzzy. Recall from Chapter 8 (The Evolutionary Approach) the connection that Eleanor Rosch makes to human reasoning. According to Rosch, human thinking is generally continuous in nature and without clear categorical boundaries. In other words, human reasoning elements are "fuzzy" and the AI systems designed on this basis seem to reflect our thinking in a direct way.

Artificial Neural Nets (ANNs)

Although Fuzzy Logic has great value and has given rise to many practical devices, scientists within AI remain frustrated because the ultimate challenge of building an automaton that behaves like a human being remains elusive. This

frustration gave rise to a third category of AI—the **Artificial Neural Net** (ANN) machine. ANN research has a long history, going back to the time when neuroscientists developed models of nervous systems. The ANN concept in its most basic form is very "simple": build a machine that replicates the human brain and just "let it think" and learn on its own. Could we teach the machine just as parents teach infants? As with other avenues of scientific investigation, numerous challenges confront ANN technology. (See the discussions of the associative cognitive model in Chapter 7.)

Overall Evaluation of the Operational Perspective

For the past several thousand years, machines that imitate life have fascinated humans; AI remains the contemporary embodiment of these endeavors and continues to stir our imaginations. While the workings of the brain are only gradually being revealed, scientists have moved forward with machines that imitate human reasoning. Among the various AI technologies, only neural net architectures—connectionist machines—have attempted even a simplistic imitation of the physiological workings of the brain. Over the past few decades commercially viable machines have emerged that have proven to be useful in assisting human enterprise in its varied forms. Of particular note are those capable of logical inference: ESs and Fuzzy Logic systems. Neural nets are particularly helpful for pattern recognition (e.g., vision systems) and have been discussed in relation to the connectionist view of cognitive science.

Despite ongoing debate about the relationship of these systems to human intelligence, development of new devices continues apace. For example, modern automobiles are being fitted with Fuzzy Logic devices that determine road conditions in real time and adjust the car's suspension system to maximize passenger comfort. Moreover, development of such machines may provide insights for cognitive scientists and provoke research into new understanding, or confirmation, of theories regarding human intelligence (Silverman, 2005). Such machines will undoubtedly create new stresses on our social and economic organizations (e.g., the impact of automation on employment). For better or worse, development of automata built on models of human intelligence is a continuing fact of life.

In Depth: The ID3 Algorithm

One of the significant challenges facing an ES is the need for a KB that is pertinent to the domain of interest. Traditionally this required a knowledge

engineer to elicit information from the expert. J. Ross Quinlan has developed a program that can elicit knowledge directly from the expert. This program—ID3 and its successors, including the latest version (C5.0), released in May 2004—was built on a concept known as **Concept Learning System**, or CLS. The goal of this program (as well as of KB building) is to develop a classification scheme so that the computer can readily respond to inquiries in an efficient and effective manner. Simply stated, ID3 is a classification algorithm in which a **decision tree** is built from a fixed set of examples. The result can be used to classify future samples. A relatively well-known ES built on such a structure is the one that banks use to decide whether or not to issue a loan to an applicant. Such a decision tree is shown in Figure 11.16.

As shown in the figure, a tree consists of branch nodes representing choices or alternatives and leaf nodes representing classifications or decisions. ID3 builds a decision tree from a fixed set of examples. The examples have several attributes and belong to a class such as yes or no. The basic CLS algorithm proceeds over a set of training instances as follows:

- If all instances are positive, then create a YES node and stop.
- If all instances are negative, create a NO node and stop.
- Otherwise, select a feature of the sample and create a decision node.

Next,

- Partition the training instances into subsets according to the values of the feature, noted above.
- Repeat the steps (recursively) on each of the subsets.

ID3 uses this algorithm, but improves on it by adding a feature-selection heuristic. ID3 searches through the attributes of the training instances and extracts the attribute that best separates the given examples.

How does ID3 decide which attribute is the best? It uses a statistical property, called **information gain**, according to the overall level of uncertainty, derived from the formula:

$$-\sum P\log_2 P$$

where P is the probability of occurrence of each attribute of a node, and the log of P is calculated to base 2.

In other words, for each attribute of a node, ID3 determines its probability of occurrence, multiplies this number by the log of the probability, and sums

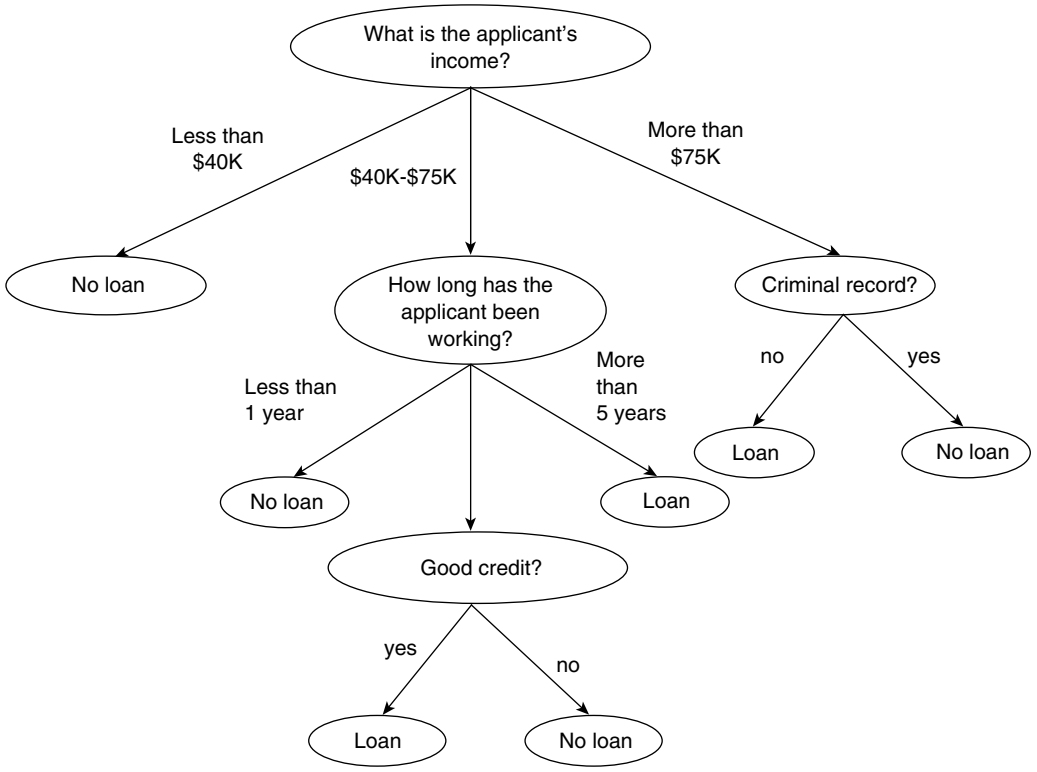


Figure 11.16 A simple loan decision tree for a banking ES

all these products to obtain a final value for the information gain. This formula can be used to determine the attribute that contains the most information—the attribute that makes the strongest contribution to the ultimate decision. Subsequent attributes are based on the possible values that the root node can have. In Figure 11.16, salary is the root node and it has three possible values: less than $40K, between $40K and $75K, and more than $75K. Each of these possible choices—children of the root node—is evaluated using the information gain formula to determine yet additional layers of the tree. This procedure continues until all attributes reach terminal (i.e., yes or no) values.

Experience with ID3 indicates that it is "user-friendly" and effective. *Its primary goal is to replace the expert* who would normally build a classification tree by hand. Information theory is the basis on which the ID3 family of decision tree induction algorithms decides on the sequence of attributes that will be used to split the data. Attributes are chosen repeatedly in this way until

a complete decision tree that classifies every input is obtained. Subsequent inquiries are automatically determined according to the tree that has been constructed. It is possible that variations in the samples can lead to misclassification, but ID3 algorithms include features that can be used to prune the decision tree with a reduction in classification errors. These algorithms learn at a reasonable rate, and the speed of the resultant decision tree classification is very high. An important aspect of ID3 is its ability to "replace the expert."

Minds On Exercise: Decision Making

Repeat the Minds On exercise found in Chapter 2 (The Philosophical Approach). Using small student groups similar to those described in that exercise, evaluate a list of ten graduate schools on the basis of: prestige, location, student-teacher ratio, library resources, and other qualifications. In the current exercise, use the techniques of Fuzzy Logic to order the resulting list.

1. How do the fuzzy choices compare to the algorithms devised with respect to solving the problem presented in Chapter 2?
2. Are any of the algorithms described for the original problem similar to those of Fuzzy Logic?
3. Do the formalisms in Fuzzy Logic implicitly reflect how humans think?

Food for Thought: Discussion Questions

1. Is it possible to develop a set of criteria to enable one to determine when to use depth-first and when to use breadth-first search methods?
2. ID3 systems can replace "experts." Are ID3 systems intelligent or simply a matter of clever programming? When does "clever programming" become "intelligence"?
3. Discuss the societal (e.g., economic, social) implications of ESs.
4. Have we subconsciously designed ESs to imitate our own cognitive processes?

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Barr, A., & Feigenbaum, E. (Eds.) (1982). *The handbook of artificial intelligence*. New York: Kaufman.,

Brooks, R. A. (2002). *Flesh and machines: How robots will change us*. New York: Vantage Books.

Jackson, P. (1999). *Introduction to expert systems* (3rd ed.). New York: Addison-Wesley.

Klir, G. J., & Folger, T. A. (1988). *Fuzzy Sets, Uncertainty and Information*. Englewood Cliffs, NJ: Prentice-Hall.

Kurzweil, R. (1999). *The age of spiritual machines: When computers exceed human intelligence*. New York: Viking Press.

McCorduck, P. (1979). *Machines who think*. San Francisco: Freeman.

Poole, D., Mackworth, A., & Goegel, R. (1998). *Computational intelligence: A logical approach*. New York: Oxford University Press.

Principe, J. C., Euliano, N. R., & Lefebvre, W. C. (2000). *Neural and Adaptive Systems*. New York: John Wiley.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hilsdale, NJ: Halsted Press.

Winston, P. H. (1993). *Artificial Intelligence* (3rd ed.). New York: Addison-Wesley.

12

Robotics: The Ultimate Intelligent Agents

"When robots finally take over the world (assuming they haven't already), we will not be able to say we weren't warned."

—A.O. Scott, New York *Times,* Friday, July 16, 2004
(in a review of the movie *I, Robot*)

Introduction

We define a **robot** here as a mechanical creature that can function autonomously. By "autonomously" we mean that the robot:

- functions without recourse to a human operator
- is able to adapt to a changing environment
- continues to function when one of its own less important parts breaks
- moves within, and changes, its world circumstances

These are truly challenging requirements for machines, and we must assume that to achieve such objectives a computer in the mechanism is required. In some real senses, the computer must have the intelligence that replicates the aptitude of a human being. Through a process of consolidation and synthesis,

Table 12.1 Highlights in Robotic History

Date	*Event*
400 B.C.E.	Archytas of Tarentum (mathematician) builds a wooden dove that flaps its wings and flies.
1500s	Humanlike creations that can play musical instruments.
1600s	Continued development of mechanisms that fly, or perform in a play.
1700s	De Vaucanson's duck that can drink, eat, and digest food.
1890	Nikola Tesla creates a remote-controlled vehicle.
Early 1900s	Development of electronic devices leading to the first electronic computers.
1942	Industrial telemanipulator that translates motions on one end into corresponding motions on the other end in order to handle dangerous (radioactive) materials.
1954–1956	George Devol and Joe Engelberger create the first industrial robot. Unimation introduces a robotic arm for industrial purposes (spray painting).
1966	Shakey, the first AI robot. Developed at Stanford Research Institute (SRI) for the Defense Advanced Research Projects Agency (DARPA). Although the device can reason about its surroundings, its limitations are obvious.
1971	The Apollo 15 mission includes a semi-autonomous explorer for extra-vehicular activity (EVA).
1973	Freddy II—demonstrates that a robot can assemble objects automatically from a heap of parts.
1975	Puma—Programmable Universal Manipulation Arm; widely used for industrial purposes.
1979	Hans Moravec develops the Stanford Cart, an autonomous vehicle that can avoid obstacles.
1984	Doug Lenat's Cyc project. A KB to help robots understand the worldview.
1989	Rodney Brooks introduces robots based on the subsumptive architecture.
1996	Michael Triantafyllou (MIT) develops a robotic fish.
1996	Honda unveils the P-2 humanoid robot that can walk, climb stairs, and carry loads.

Date	*Event*
1997	NASA's Pathfinder robot lands on Mars and carries out semi-autonomous experiments.
1998	Cynthia Breazeal (MIT) introduces the Kismet robot that mimics the emotional behavior of a baby.
1999	Sony introduces Aibo—a robotic dog.
2000	Sandro Mussa-Ivaldi (Northwestern Medical School) connects neurons within the lamprey brain to sensors in order to control a robot.
2001	Global Hawk robotic spyplane charts its own course from California to Australia.
2005	Scientists at Duke as well as Cal Tech develop technologies that will enable primates (and ultimately humans) to operate machines exclusively through brain signals. At Cal Tech, scientists have been able to decode conscious intent.

a system for designing and analyzing these machines has evolved. We will explore this paradigm in this chapter.

In Chapter 10 (Artificial Intelligence I) we noted that the desire to create an automaton in our own image has existed for a thousand years. Today, that appetite has reemerged, fired by advances in computer technology; according to some we are within sight of achieving that ultimate goal. The word "robot" was coined by Czech writer Karel Čapek in 1921, who used the word in his play, *R.U.R. (Rossum's Universal Robots)*. It comes from the Czech word meaning "menial labor." Čapek presented these machines as smart enough to replace a human in any job. In science fiction literature, robots are viewed as humanlike servants made of mechanical parts. While a mechanical humanlike form can readily be constructed, an ultimate intelligent agent incorporating the intellectual and physiological attributes of a human has not yet appeared in our culture. In recent decades, robotic researchers (**roboticists**) have therefore chosen to approach the problem from a more modest perspective. In particular, many robotic creations have taken the form of insects, dogs, and even fish, with the level of intelligence of these creations imitative of their natural models. Table 12.1 summarizes some historical highlights in robotic development.

Some Robotic Achievements

Some of the more recent achievements in robotics trace their origins to the dawn of the modern nuclear age, when humans attempted to manipulate dangerously radioactive materials in a safe manner. As with many other advances, war, or the preparation for war, has led to advances in robotics. The devices that came into being featured mechanical "extensions" of our own limbs, permitting us to manipulate material at a safe distance. They were far from autonomous, but we learned much from the design of artificial limbs that approximated human counterparts. In particular, the artificial limbs could imitate movements of joints and could transmit the "feel" of the limb movements to the human operator. Over time, design of such manipulators was refined using the concepts of control that were popularized by Norbert Wiener (Wiener, 1965).[1] The industrial community came to view such machines as highly desirable for circumstances wherein jobs were *d*irty, *d*angerous, or *d*ull—the "3D's" of robotics. Figure 12.1 sketches the timeline of robotic development, particularly as it relates to recent history.

The space exploration and exploitation projects initiated by John F. Kennedy and NASA in 1960 renewed the need for truly autonomous robots. Autonomous robotic explorers have been a keen area of research and development

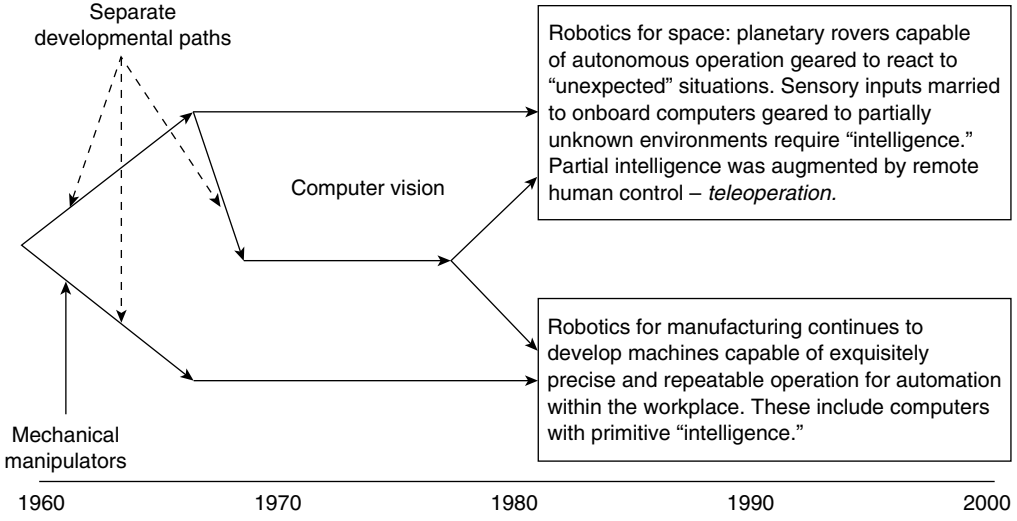

Separate developmental paths
Robotics for space: planetary rovers capable of autonomous operation geared to react to "unexpected" situations. Sensory inputs married to onboard computers geared to partially unknown environments require "intelligence." Partial intelligence was augmented by remote human control – *teleoperation.*
Computer vision
Robotics for manufacturing continues to develop machines capable of exquisitely precise and repeatable operation for automation within the workplace. These include computers with primitive "intelligence."
Mechanical manipulators
1960
1970
1980
1990
2000

Figure 12.1 Recent developmental paths in robotics

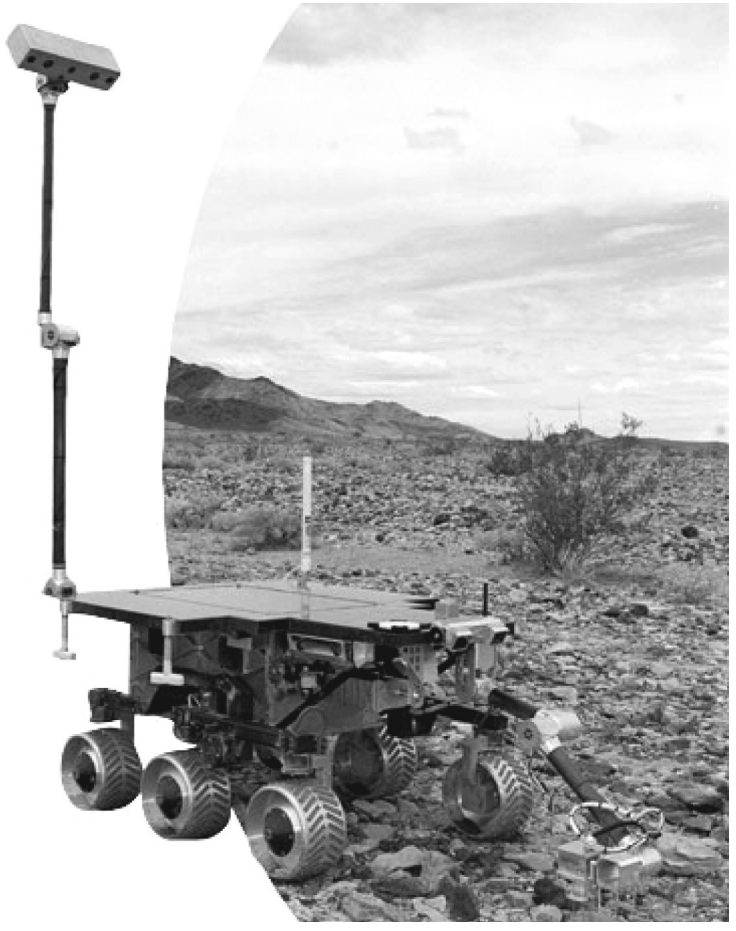


Figure 12.2 The Field Integrated Design and Operations (FIDO) rover

since the start of the Space Age. One ambitious example of this kind of robot is shown in Figure 12.2. It is representative of many autonomous robotic designs of the late twentieth and early twenty-first centuries.

FIDO is designed to sample rock and soil on a remote planet (e.g., Mars). Its capabilities include the potential analysis of planetary material, as well as intelligence for surface rendezvous with natural and manmade objects. Ultimately, such vehicles would be able to gather and return samples from remote bodies. NASA supports the Jet Propulsion Laboratory (JPL) at the California Institute of Technology.

The development of robots for industrial purposes has continued to be more concerned with such applications as providing intelligence that will adjust to circumstances within a manufacturing environment, in which parts being fabricated may not turn up in the correct orientation. Historical developments in robotics have avoided the need to understand the ultimate meaning of "intelligence." Instead, focus has been on some of the generally recognized characteristics of intelligence, such as learning, planning, reasoning, problem solving, knowledge representation, and sensing (e.g., computer vision). Recent developments have shifted the focus to the fundamental understanding of functional intelligence within the human (e.g., "How do humans react to unanticipated situations?"). In this regard we cite the work of the Artificial Intelligence Laboratory at MIT, which will be discussed later in this chapter.

Evaluating Robotic Potentials

Consider for a moment some of the aspects of human versatility that a robot would need to recreate in order to approach equivalent capabilities: it would need to understand speech, have vision capability, be able to plan and reason, represent a world model (environment), and learn. Clearly these are awesome challenges and we are only at the dawn of replicating these capabilities; our accomplishments to date would be considered crude by most robotic experts, despite some very sophisticated creations, such as the one shown in Figure 12.2.

Biological and Behavioral Foundations of Robotic Paradigms

Rodney Brooks notes that people who visit the AI Lab at MIT will interact with the various robots in very human ways (see the list of Suggested Readings). After one such visit, Professor Sherry Turkle, observing the robot named "Cog," noted the following in her book *Life on the Screen* (1995):

> Cog "noticed" me soon after I entered its room. Its head turned to follow me and I was embarrassed to note that this made me happy. I found myself competing with another visitor for its attention. At one point, I felt sure that Cog's eyes had "caught" my own. My first visit left me shaken—not by anything that Cog was able to accomplish but by my own reaction to "him." For years whenever I had heard Rodney Brooks speak about his

> robotic "creatures," I had always been careful to mentally put quotation marks around the word. But now, with Cog, I had found the quotation marks had disappeared. *Despite myself and despite my continuing skepticism about this research project, I had behaved as though in the presence of another being.* [Italics added.]

Professor Turkle's experience is not unique. Many people have a tendency to ascribe human characteristics to robots, even those that exhibit only rudimentary behaviors. We tend to treat machines as if they had intentions like our own. It can therefore be useful for us to inquire: What are the important principles that can be extracted from a consideration of the physiological and psychological origins of natural intelligence? And prior to considering the major paradigms of the robotic world, we will review some important foundations of several theories about the origins of intelligence. Physiologists have been able to visualize and precisely measure animal neural and physiological signals since the early part of the twentieth century. During the late 1970s and early 1980s, Michael Arbib began to investigate abstract models of animal intelligence, including both biological and cognitive elements, in order to gain insights into robotics (Arbib, 1981). Also worthy of historical mention is the work of Valentino Braitenberg, who proposed a series of hypothetical vehicles of increasing complexity with capabilities that "mimicked" the course of evolution in primates (Braitenberg, 1984).

Studies in biology can help us in the design of robots as well as of other machines; the biological world can teach us much. For example, all parts of an airplane's wing structure are inspired by, and imitative of, a bird's wing and the functioning of a bird's wing (except wing flapping). The study of intelligent biological agents unquestionably has value in our quest to understand how humans may function, as well as in our desire to produce a robot. Animals—even lower ones—function in an **open world** paradigm. Insects, fish, and frogs exhibit exceedingly intelligent behavior, even though they have virtually no brain. By an open world, we mean a world in which the actions of the agents alter the world and force those agents to establish a revised view of that world. Once a frog snatches a fly, that insect is no longer in the frog's visual field. An alternative paradigm for robotic applications is the **closed world** viewpoint, which assumes that the world model contains everything the robot needs to know—there are no surprises.

Our daily activities can be viewed from a variety of perspectives. Consider a very brief list of the activities of daily living: walking across a street, eating a meal, phoning a friend, and crying. Scientists have sought to explain or describe such phenomena using different models. Crossing a street might be

described in terms of the firing of nerves that, in turn, innervate muscles. If these signals are coordinated in time, our legs carry us across the street. The same activity need not be viewed with references to anything within the body. If we see a green light—a stimulus—we respond by crossing the street. Five distinct ways of studying our actions are possible: neurobiological, behavioral, cognitive, psychoanalytical, and phenomenological (humanistic). The premise of this text is that the mind processes information that it receives and transforms it into new forms, and that this processing of information is the foundation of the cognitive approach to describing our actions. One element of these processes is the relationship of brain activity to events in the peripheral nervous system and its consequent implication for our actions: this is the neurological description and has been included as appropriate in this text. (See Chapter 6, The Neuroscience Approach.) The basic assumption of the psychoanalytic approach is that our actions stem from processes that are unconscious (i.e., thoughts, fears, and wishes, of which we are unaware but which influence our actions). The phenomenological approach focuses on subjective experience or the individual's perception and interpretation of events.

Of particular importance for robotics and the potential to create an intelligent robotic agent is the behavioral approach. Robotic models attempt to simulate the observable activities of an organism—the behavioral paradigm. One of the early advocates of this approach was John B. Watson (Watson, 1919). This approach is the current focus of research in robotics. Watson's model is founded on the stimulus-response model of simulation—we see a green light (stimulus), we cross the street (response). As described below, this is an important component of robotic modeling.

From our perspective, observable biological behavior is a fundamental building block of natural intelligence. A **behavior** is considered to be a mapping (translation) of sensory inputs onto a pattern of motor actions whose purpose is to complete a given task. Behavior can be viewed from a number of perspectives. This is depicted in Figure 12.3.

The "N-gram" is a map or transformation of the sensory data into motor action. N-grams are representations of sequences of "elementary" actions that are stored in our brain, such as the transformation of notes in a musical score into our fingers' positions and actions in relation to a musical instrument. **Reflexive responses** last as long as the stimulus that produced them, and the magnitude or intensity of the response is proportional to the stimulus intensity (e.g., removing one's hand from a hot surface). Reflexive behaviors further include **fixed-action** patterns of response, wherein the response persists for a longer duration than the duration of the stimulus. Consider cases in which animals flee predators even though the danger may have ceased. **Reactive**

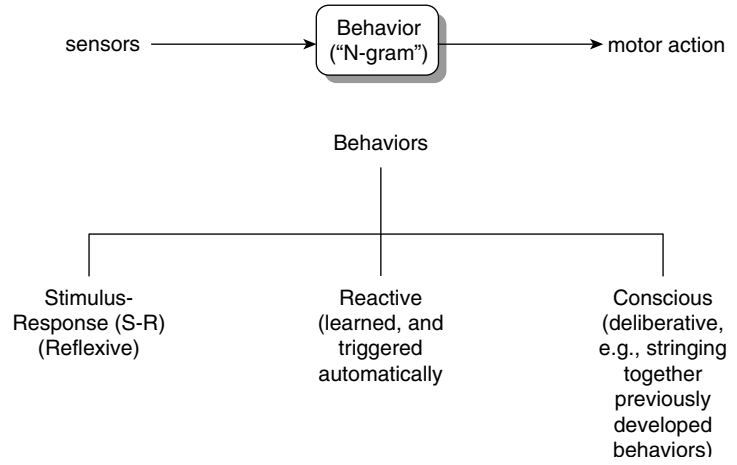


Figure 12.3 Classification of animal behaviors

responses are learned behaviors and consolidated such that they are executed without conscious thought. Subsets of such responses include **taxes**, where the organism moves to a particular orientation. For example, hatching turtles move toward the ocean using reflected moonlight. **Deliberative behavior** includes a conscious component, where previously developed behaviors may be strung together (e.g., assembling a robot). In summary, behavior is a coupling of perception and action, an integration of external and internal stimuli. The fact that we can often quantify animal behavior suggests that computer programs can be written to reproduce such activities.

The ability to quantify human behavior is a foundation for the development of computer programs that emulate intelligence. Such behavioral classifications are directly applicable to the design of robots; they can be simulated (and fabricated) and tested. Moreover, robotic demonstrations permit us to gather understanding of the advantages or deficiencies of corresponding behavioral paradigms. Cognitive psychology (human thinking and knowledge representation) and **ethology** (the study of animal behavior) are important to an understanding of, and the design of, intelligent agents such as robots. Within this book, a great many ideas about human thinking and knowledge representation have been described.

The psychologists Konrad Lorenz and Niko Tinbergen are considered to be the fathers of ethologic scientific investigation (Lorenz, 1980; Tinbergen, 1972). Using the results of their work with terns and other animals, they have proposed that there are four ways in which animals acquire behavior:

- Innate: The animal is born with a single, comprehensive behavior (e.g., feeding behavior).
- Innate, sequential: Each step in the behavior sequence is triggered by the combination of the internal state of the animal and the environment in which it is functioning. (An example would be the mating cycle of digger wasps, wherein mating is followed by nest building, which is followed by egg laying.)
- Innate with memory: A behavior is present in some form at the birth of the animal, but it is one that requires some initial stimulus to cement the behavior. (The navigational behavior of bees is an example of such behavior. The baby leaves the hive and travels short distances; these distances are gradually increased.).
- Learned: A new behavior evolves from existing behaviors. (An example is hunting, which is composed of sub-behaviors such as searching, stalking, chasing, etc.)

If we were to reproduce an intelligent agent (robot) according to the strong AI model, it would behoove us to develop an automaton along the lines suggested by Lorenz and Tinbergen. They proposed an innate reasoning mechanism (IRM) that is depicted in Figure 12.4. An IRM is similar to a reflex but operates in an open world, whereas a reflex operates in a predetermined environment.

An IRM is demonstrated when a frog reaches for its prey (e.g., a fly) with its tongue as the prey crosses its field of vision. The motor sequence (tongue

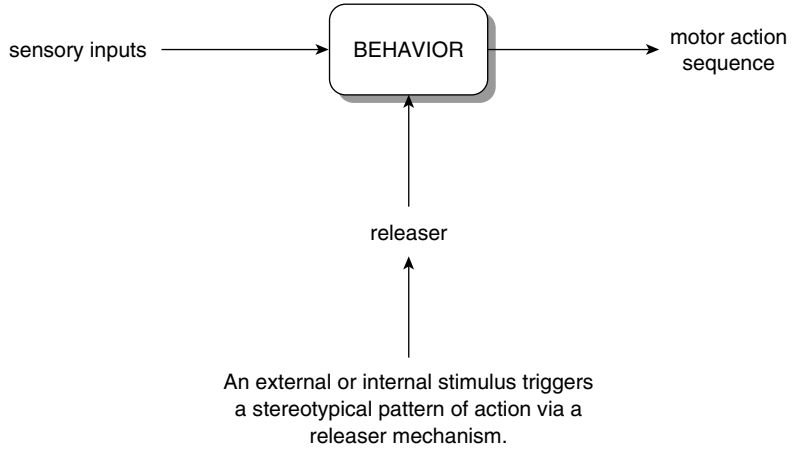

sensory inputs
BEHAVIOR
motor action sequence
releaser
An external or internal stimulus triggers a stereotypical pattern of action via a releaser mechanism.

Figure 12.4 The Lorenz/Tinbergen innate reasoning mechanism for coordination and control of behavior

action) is released by the presence of the fly (i.e., the releaser). Releasers can be compound, as when feeding occurs in an animal when food is available and the animal is hungry. In addition, behaviors can be chained (sequenced) by their releasers (Silverman & Dworkin, 1991). Simple behaviors that are independent may appear to an outside observer as a complex sequence of actions. (Rodney Brooks radically changed the design of robots using this idea [Brooks, 1999]. His contribution is explored later in this chapter.) The concurrent execution of two or more IRMs can explain such behavioral phenomena as:

- Equilibrium: circumstances in which behaviors balance each other out. An outside observer would note indecision on the part of the animal during such activity.
- Dominance: circumstances in which one IRM prevails—a kind of "winner-takes-all" phenomenon. An animal will flee from a predator regardless of where it was in a feeding behavioral sequence.
- Cancellation: in which a fight or flight situation can initiate a third behavior—the animal withdraws from both fight or flight alternatives. (Male stickleback fish build another nest when their territory happens to overlap that of a neighbor. Rather than either attack the neighbor or defend his own territory, the stickleback builds a new nest!)

Evaluation of the Lorenz/Tinbergen Approach

The models proposed by Lorenz and Tinbergen provide numerous insights into behavior. However, they fail to adequately account for the dynamic aspects of behavior. Their models reflect a "top-down" philosophy and do not sufficiently account for perception. A perception is a releaser of behavior; this is evident in an IRM. However, perception should also include guiding mechanisms. For example, a predator may trigger a flee(ing) response in an animal. But it must also trigger the animal's extraction of the information needed to accomplish the fleeing action (i.e., the safe exit). Therefore Ulric Neisser as well as J. J. Gibson—psychologists whose work led to an "ecological approach"—contributed to an alternative dynamic for describing behavior (Neisser, 1976; Gibson, 1986). Their action-perception cycle for describing animal behavior is depicted in Figure 12.5.

Neisser, using neurophysiological data, postulated two perceptual systems in animals. The first—the direct perception system—consists of structures in the more primitive regions of the brain and account for **affordances**. Affordances are perceivable potentialities within the environment that serve as triggers of an action. For example, for baby arctic terns, a red color is perceivable and

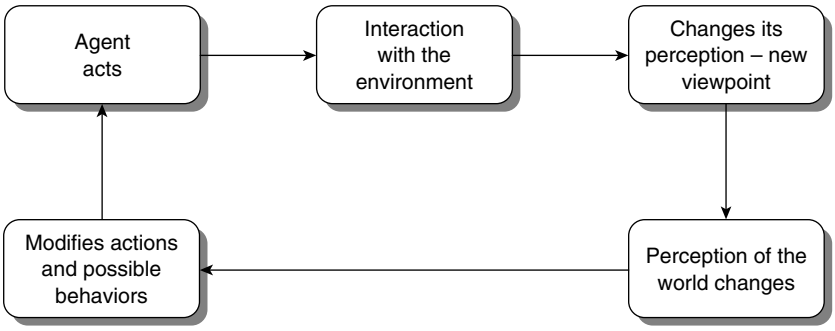

Agent acts
Interaction with the environment
Changes its perception – new viewpoint
Perception of the world changes
Modifies actions and possible behaviors

Figure 12.5 The action-perception cycle of animal behavior

represents the potential for feeding, because that is the color of the mother's beak. The second perceptual system is the recognition system. This is associated with the higher-order parts of the brain where problem solving and other cognitive activities take place. This is where top-down, model-based perception occurs. Affordances are important for roboticists because they are directly perceivable; they don't require (higher-order) memory, inference, or interpretation, and resulting actions can execute rapidly. (Additional discussion of sensory-motor perception as well as the view of Gibson are presented in Chapter 13; see, in particular, Figure 13.3.)

Transferring Behavioral Models to Robotics

Roboticists have adopted a number of behavioral concepts from physiological psychologists and their research. Robotic designs that incorporate these concepts are the basis for a true intelligent agent—machines with complex actions that are integrated formulations of (simple) independent behaviors, with tightly coupled sensing and acting. These lead to designs that operate in an inherently parallel manner and are distributed, as opposed to the intelligent agent's being a single central processing element. A robot programmed to follow a path defined by an edge could use several sensors, processes, and actuators. These might include:

Sensors: sonar (sound), camera (vision), inclinometer (detect grade)

Processes: Path following; obstacle avoidance; speed control; camera position

Actuators: drive motor; steering motor; camera pan motor

An agent should rely on Boolean activation mechanisms, characterized by the presence or absence of a stimulus, such as IRMs. The presence of an obstacle in the path of a robot will immediately trigger an avoidance process. The existence of the obstacle would be signaled by the presence of a variable such as information arriving at a vision sensor.

To simplify sensing, a perception mechanism will filter stimuli and consider only what is relevant to the behavior. This is summarized as **action-oriented perception.** For example, variation in information hitting a sensor might be due to fluctuations in light intensity. Such interferences must be eliminated if the intelligent agent is to obtain the true signal.

Direct perception (affordances) reduces the computational complexity of sensing, and can trigger action without extensive memory, logic, or reasoning. For robots with articulated legs, the stability of the mechanism would be governed by a direct stimulus-response process, rather than the transmission of this information to a central computational element that must concurrently control a number of activating motor elements.

Behaviors should be activated in an independent manner. The interaction of competing behaviors can be enhancing or inhibiting, as when a resolution of the competition is necessary for appropriate action. A robot designed to rescue people from dangerous situations would need to establish priorities in the event that there is more than one individual in danger.

Evaluation of the Biological Basis of Robotics

Unfortunately, natural intelligence is not fully understood at this point in time. What biologists and psychologists do not fully understand are clearly challenges for robotic design. Physiological psychologists cannot account for a number of phenomena that, if understood, would greatly help in the design of robots. Some elements of robotic design that are particularly difficult to achieve include: concurrent behavior conflicts, missed affordances, and the ability for the robot to learn through experience.

For example, robots may be asked to look for fire in a burning building and rescue people in a concurrent manner. How shall these be handled? Should the robots choose the dominant process? Combine these requirements? Cancel them and introduce another process? The handling of concurrent behaviors is a problem for humans and robots alike.

Direct perception and its associated sensory-action activities are efficient. In some circumstances it is not always clear when some knowledge and its associated memory requirements have been overlooked. For example, a human observer might immediately recognize a brand of soda by the color of the

container. However, a robot using its complement of sensors might perceive the angle between the container and the robot and the size of the container. In effect, the robot must perform additional operations before it can extract the key characteristic of the container (i.e., the color). The robot has missed an affordance.

Higher-order animals have the ability to generate complex behaviors through a process of learning. Understanding of these processes is not fully resolved among cognitive scientists. Until this is resolved, it will remain a significant challenge for roboticists.

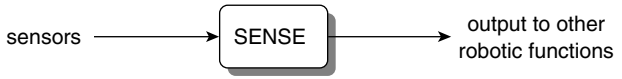
Foundations of Robotic Paradigms

A **paradigm** is a philosophy or approach for developing theories and for analyzing and evaluating a class of problems, including appropriate analytical tools and associated techniques. Each of the prominent robotic paradigms includes a series of primitive functions: **SENSE**, **PLAN**, and **ACT**. These are depicted in Figure 12.6.

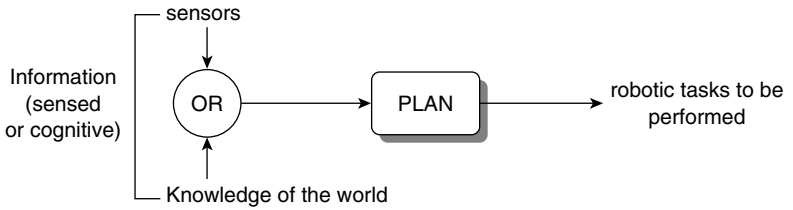
In brief, sensing includes that part of a robotic system that converts elements of an environment into information that other parts of the system use. The devices underlying this conversion are analogous to our own sensing facilities (e.g., eyes and ears). Sensing may also include some information processing features. These are summarized in Figure 12.7.

As shown in the figure, information from the sensing units may be "primitive" (i.e., "raw"), meaning that no further processing occurs. In such instances, whatever element receives such information may bring to bear its own methods to the data. Alternatively, the sensor may contain additional "intelligence" and produce data representing a more complete view of the environment—the global or world model extracted from the raw data. (Some animals have primitive visual systems, but the output from their visual sensors together with some local neuronal interconnections, in fact, reflect the world model that they need for survival.)

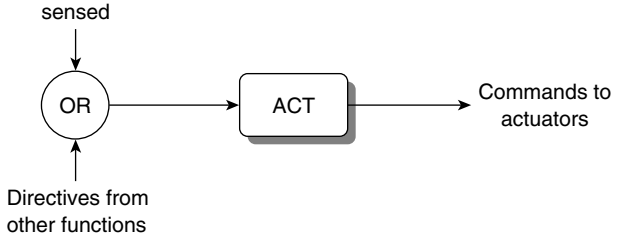
Planning elements (see Figure 12.6) correspond to one model of our own cognitive and reasoning capabilities, and actuating elements are consistent with our motor elements (e.g., muscles and the associated afferent neural structures). The fundamental robotic elements are intended to emulate the fundamental cognitive as well as the biological processes of humans. It is theorized that by interconnecting several robotic systems based on these foundation elements, it may be possible to recreate an intelligent agent that rivals human beings.



Sensing primitive: a function that takes in information from a robot's sensors (e.g., sound) and sends useful output that can be used by other functions (e.g., the direction of the sound)



Planning primitive: takes in information from a sensor or from a KB (e.g., a plan of the environment in which the robot is to function) and generates a set of tasks for the robot to execute (e.g., go down the hall for 10 meters and turn left).



Act primitive: produces commands to the motivating (e.g., wheels) or other actuators (e.g., a microphone boom) of the robot.

Figure 12.6 The primitive robotic functions

Evaluation of Paradigm Foundations

Most roboticists (as well as cognitive scientists) accept the sense, plan, and act "primitives" of robotic paradigms as reflecting elements of human intelligence. As such, they would constitute corresponding processes within an intelligent (machine) agent. As noted above, the ability to learn is a biological feature of more advanced animals. A growing number of roboticists believe that a new primitive needs to be added to robotic architectures, namely a

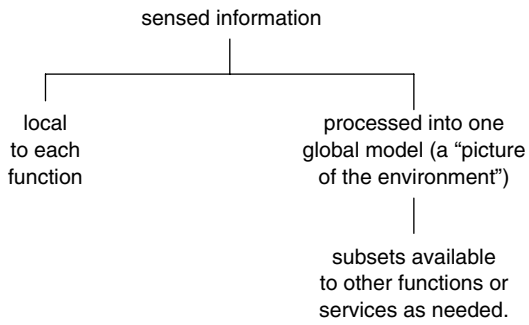


Figure 12.7 Sensory information processing in robotic systems

learn(ing) process. However, there are presently no formal architectures in which a learn(ing) process has been well integrated. When this happens, there will surely be a paradigm shift and an accompanying expansion of the types of automata within the literature, as well as within the public domain.

Robotic Paradigms

Robotic paradigms seek to replicate the behaviors identified within the animal world and fall into one of three categories: the Hierarchical Paradigm, the Reactive Paradigm, and the Hybrid Paradigm.

Hierarchical Paradigm

Also referred to as the "top-down" approach to robotic design, this paradigm reflects how humans include "planning" as a key element in their completion of given tasks. Robotic models that used this paradigm received a good deal of attention from about 1967 to 1990. (Murphy, 2000). The basic architecture is shown in Figure 12.8.

This top-down paradigm is highly sequential. The robot senses the world and constructs a map—a global view—of the world, including everything that its sensors can detect. Sensing, in this instance, thus includes additional processing, as seen in Figure 12.7. This worldview, together with additional information that may be stored in the robot's (computer) memory, is used to create a plan by which the robot intends to reach its goal. Information within the memory may include cognitive rules, for example, knowing that to go from

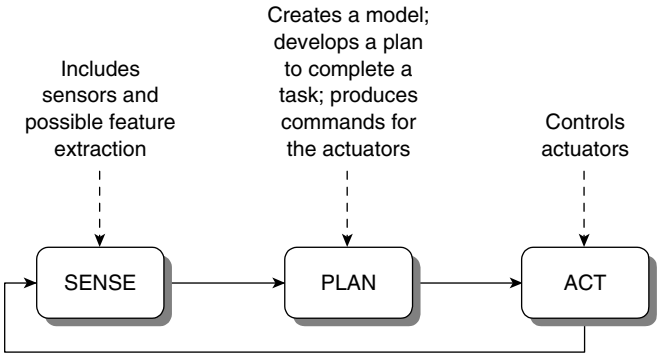

Includes sensors and possible feature extraction
Creates a model; develops a plan to complete a task; produces commands for the actuators
Controls actuators
SENSE
PLAN
ACT

Figure 12.8 Basic architecture of the Hierarchical Paradigm

one room to another may require an open door. The plan(ing) algorithm will send a series of directives to any onboard robotic actuators (e.g., a motor). The robot will proceed to carry out the first directive, which may be a "complex directive" that reflects a series of subgoals. With each directive, the robot reevaluates the worldview. This refined information is then recombined with the system's cognitive elements to reconfigure the plan, producing yet another, or altered, set of actuator directives. This sequential ("loop") organization continues until the robot completes its task.

The classical robot designed according to this principle was named "Shakey." (It got its name from the type of motion that it exhibited.) It was developed at the Stanford Artificial Intelligence Laboratory (SAIL). The algorithm that was employed, called **Strips**, was a variant of the General Problem Solver (GPS). (See Chapter 5 for a discussion of GPS.) Strips uses means-ends analysis: if the robot's goal cannot be achieved by a single directive, the planner chooses a sub-task that will reduce the difference between the current state of the robot and the goal state. For example, if the robot is located in California and it must be able to end up in New York, one could envision the distance between location and target (i.e., 3,000 miles) as the "difference" parameter to be reduced. The program could then decide that flying was the best method to reduce this difference. This choice would be included in its cognitive KB. The planner would then develop a series of subgoals to complete the flight.

Evaluation of the Hierarchical Paradigm

The hierarchical paradigm would seem to be a simple, straightforward representation of how humans function. Within the hierarchical organization,

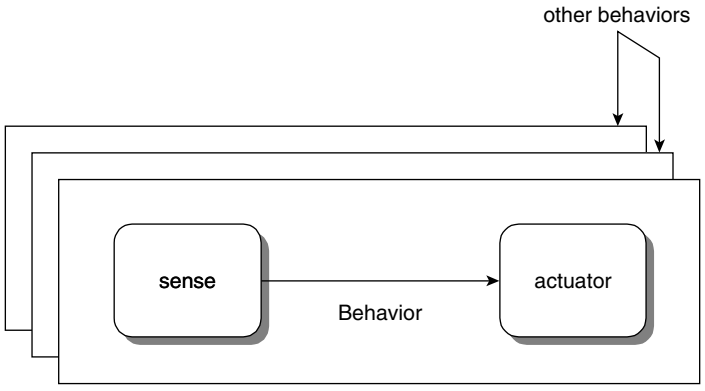
PLAN was to reflect the way people "think." However, do people really (or usually) "think" about some action in a "conscious" sense? They may, for example, have a repertoire of default schemes for completing a task. Consider the "N-gram" model noted in Figure 12.3. (A skilled musician does not have to "stop and think" about each note in a score. Physiologically speaking, it would not be possible to see a note, and, using the position of the note on the instrument, develop a plan for finger action, and then execute the action in a timely manner. There must be a more direct path between the printed score and the finger [or voice] action.). The Hierarchical Paradigm in Figure 12.8 lays out robotic (as well as human) tasks as a horizontal decomposition (Brooks, 1986). Ethological literature describes behavior as a vertical organization in which an agent starts with primitive survival behaviors and builds upon such facilities to develop new ones. The building process may reuse primitive behaviors, may inhibit such behaviors, or create parallel tracks of more advanced behaviors. The resultant functional architecture of this arrangement has a vertical (parallel) appearance and has given rise to the Reactive Paradigm.

An additional limitation of the hierarchical model derives from its single global model of the world. Generic global world models do not handle "surprises" very well. For example, if the position of the target has changed or unexpected obstacles were to appear, a great many prior directives would be voided. Thus, Shakey, as well as the Hierarchical Paradigm itself, encountered the frame problem head on. (See Chapter 11 for a discussion of frames.) A static worldview does not lend itself to adaptability easily. Moreover, the closed world assumption precludes the possibility of the dynamic circumstances usually encountered by humans. The rise of the Reactive Model significantly reduced interest in the Hierarchical approach.

The Reactive Paradigm

The Reactive Paradigm has developed from the idea that complex behaviors grow out of simple behaviors that operate in a concurrent manner. For example, Brooks built insect-like robots using a combination of simple sense-action processes. (This approach is known as **subsumption**.) This idea is illustrated in Figure 12.9.

In developing this concept, Rodney Brooks actually rejected a principal element of AI; he eliminated the "reasoning" process. There were to be no chains of thought in his robots. He would build robots with "unthinking activity." There would be direct connections between perception (sensory information) and action. He decomposed robotic behavior into a series of simple elements.



Complex behaviors are developed by control systems that coordinate the various behaviors.

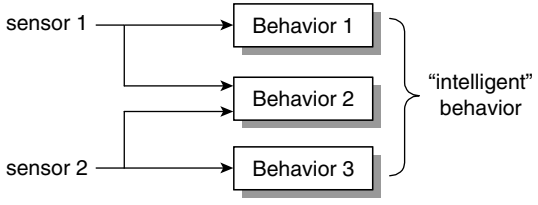


Figure 12.9 Vertical organization of the Reactive Paradigm for a simple robot, after Rodney Brooks

These would be implemented in low cost processing units that would take a few milliseconds (1 millisecond = 0.001 second) to complete a computation. But certainly, behavior is far more complex than the simple operations managed by these processors. To develop robots capable of more intelligent behaviors, he added control systems that made use of the simple, primitive capabilities of the older microprocessors. The control systems used appropriate sensors to invoke an appropriate combination of behaviors, as shown in Figure 12.9. (This model "mimics" biological systems in which "complex capabilities are built on top of simpler capabilities, often through the addition of new neurons." See Chapter 4 for the cognitive models of the brain.) In 1985 he devised a robot with three behavioral layers. The first control system would include collision avoidance. Using sonar (sound) detectors, the robot would "shy away" from objects that it detected, even if they were moving. A second control system caused the robot to simply wander about without any serious purpose, as if it were exploring. The third control system was designed for

purposeful exploration. The robot—Allen, named in honor of Allen Newell, one of the originators of AI—could do remarkable things at the time primarily because of the speed at which it could travel (half a meter per second). It had no internal description of the world, hence it could not PLAN; it used the world as its own model. It could move down a corridor without colliding with walls, and it went around oncoming pedestrians. This became the platform for many other robots that included more and more capabilities (e.g., picking up objects).

The characteristics and connotations of reactive behaviors include:

- Speed: Rapid execution because of the direct connection ("tight coupling") between the SENSE and ACT primitives.
- Simplicity: Absence of "memory" in the traditional sense. There is a direct connection between the SENSE and ACT primitives. (It should be noted that one could consider the "permanent" connection between SENSE and ACT to constitute a type of memory; it is, however, not easily alterable, as in the computer, as shown in Chapter 10.)
- "Small world": Such reactive agents are situated in an ecological niche formed by the goals, the world in which they operate, and their perception of the world. They are an integral part of the world and their actions change their world. They receive feedback from the changed world and generate a new cycle of actions.
- Simple programming: Constructing a robotic system under the Reactive Paradigm is achieved by programming a particular behavior. The fundamental component of an implementation is a behavior. *All actions are accomplished through behaviors.*
- Built on choice: Reactive architectures provide mechanisms for triggering behaviors and must determine what happens when multiple behaviors are active at the same time. (For example, the robot might have to choose between "fleeing" and simple "avoidance.") The rules for combining behaviors are often heuristic (ad hoc) and may include Fuzzy Logic methods (see Chapter 11) or "winner-take-all" algorithms.

The Reactive Paradigm described here reflects Brooks's **Subsumptive Architecture**. An alternative scheme involves potential field concepts, where sensors find "attractive" pathways to some goal. (An example of this can be seen when water flows around a rock in a stream; the flow is disturbed and this disturbance can be detected.) However, a complete description of this somewhat technical subject is beyond the scope of this text.

The Colorado School of Mines developed a small, unmanned vehicle about the size of a golf cart that had to navigate around an outdoor course of white lines painted on grass. The design won first prize in a 1994 Unmanned Ground

Robotics Competition. A key feature of this design was the effective use of a few behaviors, incrementally developed, and the use of affordances combined with an understanding of the ecological niche. The vehicle could follow a path that had hairpin turns and avoid stationary obstacles in its path as well as a sand pit. Using a simple algorithm that enabled it to follow the white line, the robot was able to traverse the course despite some unforeseen mishaps. This case provides an archetypical example among purely reactive motor schema to assign one sensor per behavior. That is, one optical sensor could detect the white line that produced a "following" action (schema) for the motors, while a second sensor signaled the presence of an obstacle (by "bumping" it), which triggered a second motor action (schema) that "went around" the obstacle.

Evaluation of the Reactive Paradigm

The Reactive Paradigm lends itself to those situations in which tasks can be decomposed into reflexive behaviors. Direct perception can be readily committed to hardware for such niche circumstances as obstacle avoidance. Whether such architectures can be ported (reused) to new applications is an open question. For example, they are not easily transferred to domains where planning or reasoning about resource allocation is essential. Navigation in a hallway is not the same as negotiating a terrain with an ill-defined geography. At the very least, some modifications of the SENSE-ACT primitives would be needed.

An additional limitation emanates from the genuine lack of robustness in such systems. Because of the necessity of reducing the complexity of such systems, they generally do not have any redundancy (e.g., a second or backup sensing system). If a sensor fails there is generally no backup available for this device and it also fails. If the software process (program) that coordinates the SENSE-ACT primitives suffers some form of degradation (e.g., a problem in the program that shows up when it encounters data that were not previously tested), the robot may still function at a reduced level using its primitive operations for survival. This is described as "graceful degradation." But the Reactive Paradigm has no mechanism for detection of such degradation (e.g., if the robot has a malfunctioning joint). Humans can adjust for deficits in or insults to brain functioning in a far more effective manner.

The detailed design of Reactive robots is an art minus the well-defined basis used in the finite state machine architectures, in which the programs detect the current state of the system and are programmed to carry out the next operation based on what is (hardwired) in their memory. The sequences of steps within Reactive robots are generally implementations of very simple ad hoc rules of behavior without the benefit of large amounts of memory. The assemblages

of behaviors depend heavily on the programmer. Designs of robots that are truly reflective of human intelligence should be "smart enough" to be able to select the necessary behaviors to achieve a particular goal and to generate an appropriate execution sequence.

The Hybrid Deliberative/Reactive Paradigm

Robots that evolved under the Reactive Paradigm (up to the end of the 1980s) had no provisions for planning optimal trajectories (path planning), making maps, monitoring their own performance (learning), or selecting the best behaviors for accomplishing a task (general planning). It was thus desirable to retain the benefits inherent in the reflexive character of the Reactive designs while (re)introducing planning and deliberation as important elements of robotic design. In the late 1980s, Ron Arkin and his colleagues introduced cognitive functions into a behavioral system that essentially emulated the evolution of intelligence (Arkin et al., 1987).

One starting point for the Hybrid Deliberative/Reactive Paradigm is the bottom-up architecture that characterizes the Reactive Paradigm. (Bottom-up organizations were described in Chapter 10—see, in particular, Figure 10.5.) Upon this structure roboticists have superimposed one or more cognitive layers for planning. This school of design has been called reactive planning. Another approach employs a design more reflective of the top-down, hierarchical methodology. Regardless of the approach, designs that use a combination of reactive behaviors and planning or other cognitive components have come to be known as the Hybrid Deliberative/Reactive Paradigm. (It will be referred to as the **Hybrid Paradigm** in this chapter.) Initially, the Hybrid Paradigm did not have much to recommend itself. The advice that a researcher might receive went something like:

- If the agent was to be designed to operate in an environment that was easy to model, the Hierarchical Paradigm was the organization of choice.
- Alternatively, if the environment was unstructured, the Reactive Paradigm was recommended.

Hybrids were originally thought to be inferior approaches because they were viewed as marrying the incompatible features of the Hierarchical and Reactive worlds—as employing the fast execution of the Reactive world with the latencies inherent in the Hierarchical architecture. Hybrid agents have been able to resolve these seeming conflicting characteristics in good part because of changes in software organizations. In particular, processing techniques known

as multi-tasking and threading permit deliberative functions to execute independently of reactive behaviors.[2] Planning activities can compute the next navigational goal at a slow pace—for example, at the same time that the current goal is being rapidly achieved—using a Reactive philosophy. In addition, software has become a highly modular enterprise, such that self-sufficient program modules (objects) can be easily mixed or matched as circumstances require.

Hybrid agents are characterized by the sequence of operations shown in Figure 12.10.

The PLAN component is understood to include deliberative processes and worldview modeling as well as task formulation. Using a global worldview the mission strategy would be generated and a set of behavior instructions for the SENSE-ACT elements issued. These would be executed until the goal is achieved or until a new set of behaviors was required. Planning combines long-range considerations with knowledge of the world that may be found in a KB. The PLAN component is considered to execute independently of the real-time SENSE-ACT modules.

Combining a deliberative and reactive element has led to a proliferation of Hybrid examples and arrangements. Thus, generalizations about such architectures do not readily lend themselves to easy encapsulation. However, some common characteristics can be noted.

- In the Hybrid model behavior includes not just reflexive behaviors but also innate and learned behaviors. This may be termed a **skill set**.
- Hybrid implementations use assemblages of behaviors sequenced over time rather than primitive behaviors.
- Hybrid models normally include a **deliberative portion** within its architecture comprised of **Path planning and Map making**.
- Hybrid models may include a **Behavioral Management** component that determines the behaviors to use as well as the order in which they are to be used.
- Hybrid models may include **Performance Monitoring**: to see if the robot is actually making progress, diagnose deficits (e.g., the robot is stuck in the mud; a sensor has failed). Other sensors may be used to corroborate these observations.

In order to classify the variety of Hybrid architectures, three attributes might be considered:

What is the deliberative portion vs. the reactive portion?

What is the architecture of the deliberative portion?

How does overall behavior manifest itself?

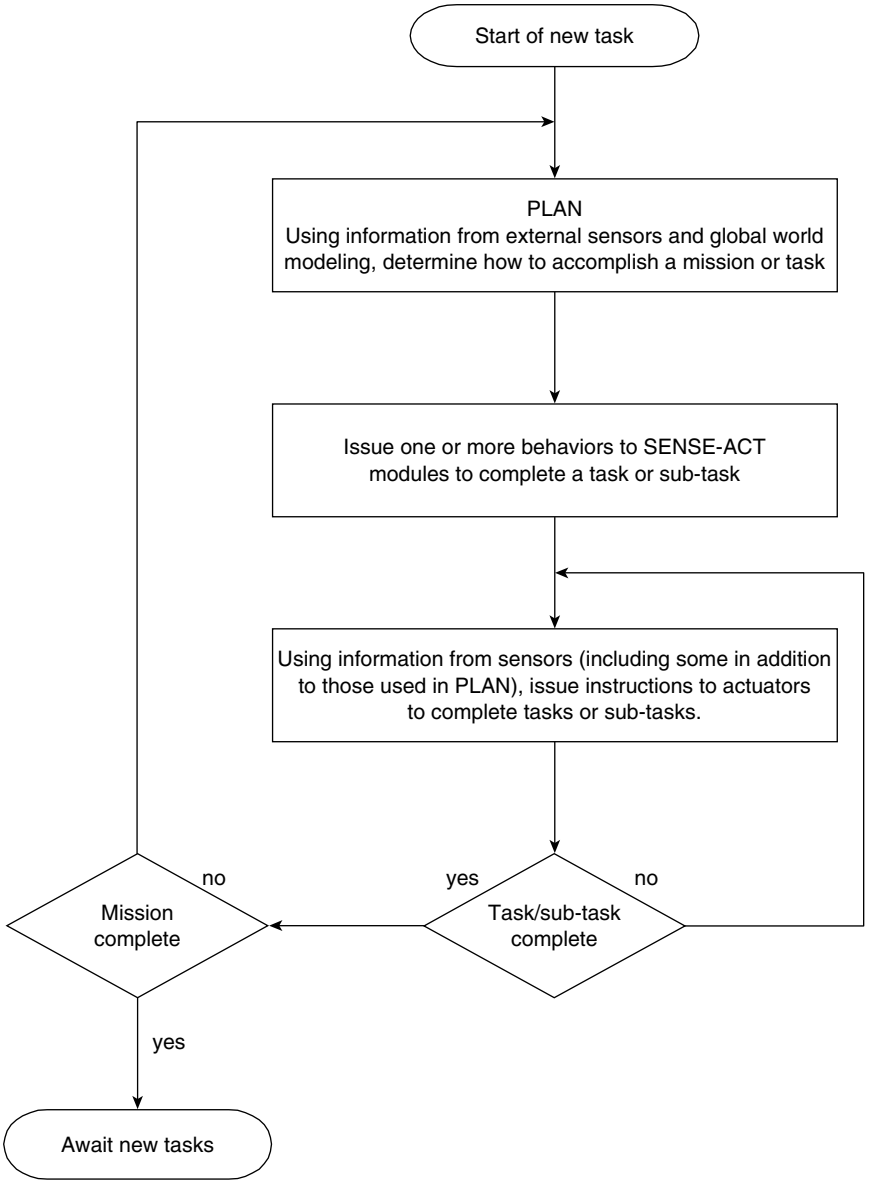

Start of new task
PLAN
Using information from external sensors and global world modeling, determine how to accomplish a mission or task
Issue one or more behaviors to SENSE-ACT modules to complete a task or sub-task
Using information from sensors (including some in addition to those used in PLAN), issue instructions to actuators to complete tasks or sub-tasks.
Task/sub-task complete
yes
no
Mission complete
no
yes
Await new tasks

Figure 12.10 The PLAN, SENSE-ACT operations of Hybrid agents

Hybrid architectures include a number of variations within the Reactive segment. These are beyond the focus of this text but are simply mentioned here without elaboration for completeness: Voting organization; Fuzzy responses; Filtering.

Table 12.2 Attributes of Hybrid Architectures

Hybrid Attribute	*Explanation*
Mission Planner	Interacts with humans and translates commands into robotic terms; constructs a mission plan.
Cartographer	Responsible for creating, storing, and maintaining maps or spatial information about the robot's environment or task domain.
Sequencer Agent	Generates the set of behaviors to use in order to accomplish a subtask.
Resource Manager	Allocates resources to behaviors—e.g., a robot may have stereo vision, sonar(s), infrared (IR) sensors, all to detect range. The Resource Manager determines the circumstances in which to use each. (In Reactive architectures, resources for a behavior are often hardwired.)
Performance Monitor and Problem Solver	Allows the robot to notice if it is making progress and/or to adjust for hardware problems within itself.

Table 12.2 summarizes functional components that may be found in Hybrid architectures.

Additionally, organizational styles of Hybrid systems fall into one of three categories, as shown in Table 12.3.

Evaluation of Hybrid Architectures

It is difficult to evaluate Hybrid organizations because systems are still evolving and the deliberative component is expanding even as this text is being written. Even though individual instances may not lend themselves to criticism, an evaluation scheme is possible.

Consider modularity. Each is highly modular and most are divided into layers, which are then subdivided into modules. As the software agent programming for AI becomes more popular, more architectures will implement deliberative modules devised by independent specialists. Currently, none of the architectural styles is particularly superior. Each must be considered in light of the application for which it is used.

Hybrids tend to have a high degree of niche targetability. Hybrids can be used for applications not appropriate for purely Reactive systems.

Table 12.3 Organizational Styles of Hybrid Robotic Systems

Style	*Explanation*
Managerial style	Subdivides the deliberative portion into layers based on the managerial responsibility of each deliberative portion (e.g., a Mission Planning element could be responsible for navigation).
State hierarchies	Uses the knowledge of the robot's state to distinguish between reactive and deliberative activities: Reactive functions have no state or self-awareness; they function only in the present. Deliberative functions require knowledge of the past (e.g., where the robot has been) or knowledge of the future (e.g., mission and path planning).
Model-oriented	Characterized by behaviors that have access to portions of a world model; these architectures are close to purely Hierarchical robots.

Consider robustness. Many designs attempt to make provision for monitoring performance and to replace or adapt reactive behaviors as needed.

On a global level, one has to ask the questions: "Are Hybrid designs really unique?" "Are they merely variations of Hierarchical architectures?" "Is the only difference between them 'how they achieve the same ends'?" Hybrids explicitly take advantage of the principles of software engineering. The two paradigms reflect different attitudes toward world modeling.

In Hybrid systems, global models are used for symbolic functions. The frame problem does not exist for the Hybrid because:

1. Execution is reactive and therefore well-suited for unstructured environments.
2. Software agents can use agent-specific abstractions to exploit the structure of an environment in order to fulfill their particular role in deliberation.

Global models reflect a closed world, but closed at the deliberative level. The robot can think in terms of a closed world, while it acts in an open world.

Another basis of comparison between Hybrid and Hierarchical models concerns the difference in the roles played by planning and execution. Shakey—the embodiment of the Hierarchical model—planned every move down to the lowest level of granularity and had problems confirming that an action had been

accomplished. Modern planners often produce only partial plans, then execute that part of the plan, note the results, and plan the next step. It is not clear whether a SENSE, PLAN, ACT sequence (with a more suitable planner) would produce more intuitively appealing robot architectures.

The final distinction may rest on the influence of biology. The Hybrid Paradigm has its roots in ethology, and provides a framework for exploring cognitive science. The Hierarchical Paradigm is less clearly cognitively plausible, although both share the same cognitive motivation. Both may suffer from limitations in computing capacity and an associated paucity of planning intelligence. In one instance, as an autonomous rover approached an unanticipated boulder, its wheels first turned sharply to the right (in an avoidance behavior), then turned sharply to the left, and finally drove the vehicle directly forward straight into the boulder! In this case the robot resolved behavioral conflict by mutual cancellation and selection of a new behavior. Unfortunately, this behavior was inappropriate.

Overall Evaluation of Robots as Ultimate Intelligent Agents

Rodney Brooks, Director of the AI Laboratory at MIT, suggests the possibility that within twenty years the computational power of the PC will surpass that of a human brain. Taking this into account (as well as the expanding use of the surgical procedures for embedding silicon and steel inside human beings to compensate for lost capabilities), one can question our mortality as well as what it means to be human. Just as it took centuries for people to accept the fact that our universe is not earth-centered, as well as the theory of evolution (and animal intelligence that is comparable to ours, in many ways), the equivalence of man and machine may some day be accepted. When we look back, the PLAN-SENSE-ACT architectures (and their variations) will seem primitive, to say the least. Consider some recent developments in robotics that do not conform to any of the historical architectures. At Georgia Tech, researchers seek to create computing systems that perform more like the human brain. To this end, they have developed a small robot that moves about using the brain signals of a rat—they call this device the Hybrot. It is a hybrid of living and robotic parts that connects laboratory cultures containing living neurons to computers in order to create a simulated animal. A droplet containing a few thousand living neurons from rat cortex is placed on a petri dish instrumented with an array of 60 microelectrodes. The neurons can be kept alive in a specially designed incubator for up to two years. The neural network is connected to a body in the form of a mobile robot. The robot moves under the command of the neural

activity that it receives, and information from the sensors is sent back to the cultured net as electrical stimuli—it is a kind of neural net that has been previously discussed. The Georgia Tech group hopes that some day these neural interfacing techniques can help people to control prosthetic limbs directly from their brain. In yet another application, scientists have been able to decode intended hand trajectories from motor cortical neurons and, using these signals, have been able to control external devices (Musallam et al., 2004).

The field of AI robotics has produced a large number of examples of machines with physical properties and navigational ability that are primitive resemblances of human behaviors. We find examples in health care as well as entertainment. Originally conceived to perform dirty, dull, and dangerous tasks, the machines are now viewed as personal assistants. A question that pervades research is: "Will such machines become truly intelligent agents in our lifetime?" There is no definitive answer to this intriguing question, but the following quotations from two individuals with the same affiliation summarize the alternatives.

> *"The body, this mass of biomolecules, is a machine that acts according to a set of specific rules . . . I believe myself and my children all to be mere machines."*
>
> Rodney Brooks,
> Director of the MIT AI Laboratory

> *"The reason there are no humanlike robots is not that the very idea of a mechanical mind is misguided. It is that the engineering problems that we humans solve as we see and walk and plan and make it through the day are far more challenging than landing on the moon or sequencing the human genome. Nature, once again, has found ingenious solutions that human engineers cannot yet duplicate."*
>
> Steven Pinker, Director of the Center
> for Cognitive Neuroscience at MIT

In Depth: Autonomous Robot Architecture (AuRA)

AuRA traces its origins to 1988, at which time Ron Arkin published his research on how to add more cognitive functions to a behavioral system

(Arkin, 1998). His work included the first of the type of robot design that used Hybrid architecture, and was conceived during his tenure at the Georgia Institute of Technology. AuRA is based on schema theory. A **schema** is a template for doing some activity. For example, there may be a bike riding schema that includes the data one would need to ride the bicycle, as well as the methods for riding a bicycle. It is a template because a person can ride different bicycles without starting the learning process from the beginning. The type of bicycle, the height of the seat, and the position of the handlebars can be supplied at the time that we wish to create a particular instance (case) of a bicycle. A behavior can be viewed as a schema that is composed of a motor schema and a perceptual schema. The motor schema is the template for physical activity; the perceptual schema embodies the sensing. Releasers would then activate the behavior. As regards human experience, vision is the default perceptual schema to navigate out of a room and the navigation process is the motor schema.

There are five identifiable subsystems within AuRA, as indicated in Figure 12.11.

AuRA, like other Hybrid systems, inherits its deliberative elements from the Hierarchical Paradigm. PLAN elements in the Hierarchical Paradigm are decomposed into three different categories that reflect how a human agent might devise a plan. In the Hierarchical Paradigm these elements are Mission Planner, Navigator, and Pilot. These functions are summarized below.

The Mission Planner receives a mission assignment from a human agent; it may also generate a mission on its own. For example, the mission might be to pick up rocks from a site on the moon or other planetary body, or perhaps to retrieve some suspicious object. It has the responsibility for translating the mission into terms that other parts of the software system can understand. It might, for example, generate a symbol for a particular rock as follows: "rock = R1." Using a map of the terrain, the Navigator generates a path from the robot's current location to its destination (e.g., the return vehicle, if this is a planetary mission to recover minerals). The Navigator creates a set of points arranged to create a (theoretically) straight line that the robot is to follow; the Pilot receives these path segments. Using this information, the Pilot determines which actions the robot must take. It might, for example, require the robot to turn 30 degrees to the left on the next leg of its path. Suppose however that the robot encounters an unexpected obstacle. In this case, the low-level (detailed) Reactive elements are able to adjust for this unanticipated occurrence.

The Cartographer contains all the information needed for navigation and corresponds to the KB of the Hierarchical model. The Pilot sends the Motor Schema Manager a list of the behaviors that are essential to the successful completion of the current part of the navigational assignment. Using schemas

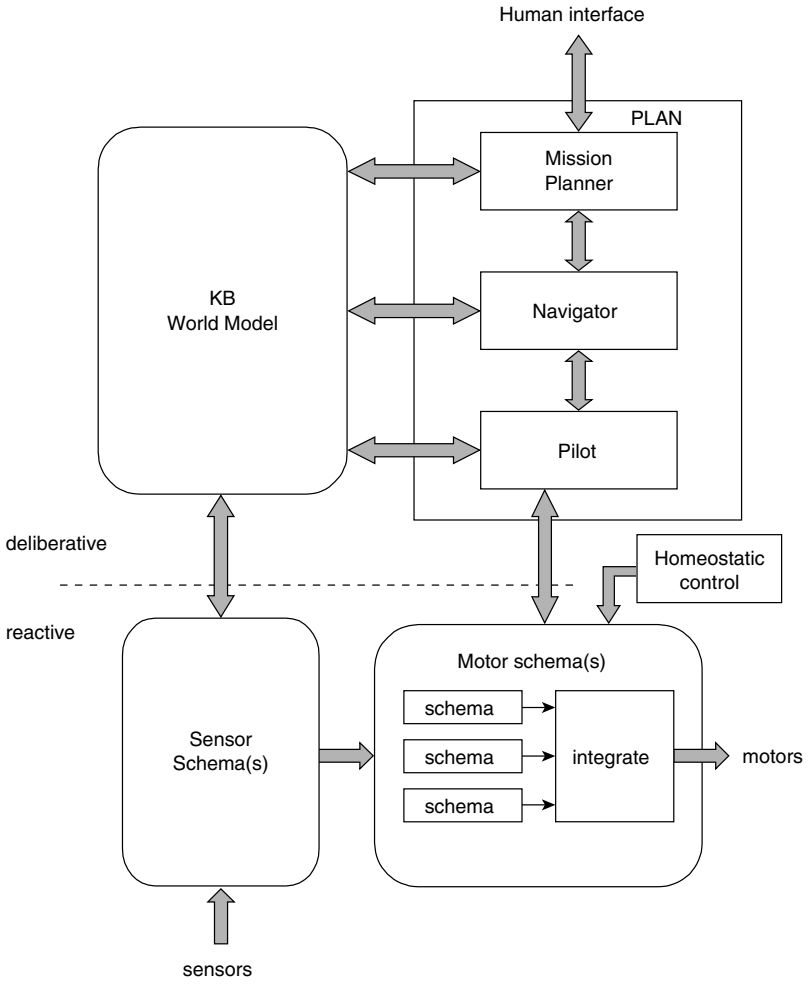


Figure 12.11 Schematic architecture of the AuRA Hybrid System

within the sensing system, the actuators will be instructed to respond with the appropriate action.

One final component of AuRA is the Homeostatic Control, which is part deliberative and part reactive. It must oversee the current state of the robot and make appropriate changes in the operation. For example, suppose that the robot has spent too much time carrying out a task and that an external deadline is rapidly approaching (e.g., the return vehicle must leave the planet at a predetermined point). The Homeostatic Control could override the instructions provided by the Pilot. Homeostatic Control approximates biological behavior; animals modify their behaviors on the basis of internal conditions.

Increasing hunger is a great releaser for increasing attention to food at the exclusion of other behaviors.

Evaluation of AuRA

AuRA, like other robotic systems, may be less suitable to a particular application or mission. (Robots have their specialized "talents" as do humans.) AuRA, in its original configuration, did not make provision for dealing with sensor failures. AuRA is representative of the managerial style of design. As an alternative, state hierarchies use knowledge of the robot's state; these distinguish between reactive and deliberative activities. Reactive behaviors are considered to act only with current information; these behaviors have no awareness of the robot's past or about the future (mission and path). This leads to a layered organization of design that employs a Planner, a Sequencer, and a Skill Manager in one instantiation of the State-Hierarchy architecture. These architectures have advantages (e.g., speed) in cases such as planetary rovers, underwater vehicles, and robotic assistants for astronauts.

Minds On Exercise: Relational Graphs

A relational graph seeks to represent the world as a graph or network of nodes and edges. Nodes represent gateways, landmarks, or goals. Edges represent a navigational path between two nodes. (The two nodes have a spatial relationship.) Such graphs can be very useful for the Navigator process within a Planner. Additional information can be attached to the edges, such as direction (e.g., North), approximate distance, terrain type, or behaviors needed to navigate that path. Build a relational graph representation that labels the distinctive places and local control strategies using gateways for a floor of a building on campus. A gateway represents an opportunity for a robot to change its overall direction of navigation. For example, an intersection of two hallways is a gateway because the robot can choose to go straight ahead, turn to the right, or make a left turn.

Food for Thought

1. Social and moral questions raised by robotics would include: Is it acceptable to use robots to fill roles in instances in which it is hard to get workers at any price? Where do we draw the line? Discuss the moral implications of the accidental death of an animal caused by a robot as opposed to a human.

2. Can ethology (animal behavior) and cognitive psychology (human knowledge representation and thinking) be integrated and applied to robotics? Provide arguments for both possible answers.
3. How would the Mission Planner of an intelligent agent construct a mission plan to "Go and bring the sheriff to me"?
4. Do the Performance Monitor and Problem Solver imply some kind of robotic self-awareness?
5. Are humans nothing more than very clever machines (robots)? Can we build a robot that will eventually surpass humans in intelligence? Will such a machine pass the Turing Test? What are the implications of recent research in which animal intentions have been detected and used to control equipment?
6. Consider how a collection of robots—a team of intelligent agents—can be used to complete a task. As an example, discuss this within the context of a series of robots that might be used for surveillance.
7. Should robotic behavior be hardwired (preprogrammed), or should robots include a process for learning from elemental procedures?

Notes

1. Wiener suggested that human communication should be the model for human-machine and machine-machine interactions. His theory—known as **cybernetics**, from the Greek word for "steersman"—was intended to improve the quality of our lives within the technological world, which is characterized by its reliance on machines and where interactions with machines are the norm.

2. Multi-tasking is a programming technique for sharing a single computer processor among several independent software tasks or jobs. Multi-threading is similar to multi-tasking, but with low processing time and with limited protection (independence), because all threads or program segments share the same memory.

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Readings

Aylett, R. (2002). *Robots: Bringing intelligent machines to life*. Hauppauge, NY: Barron's.

Brooks, R. (2002). *Flesh and machines: How robots will change us*. New York: Vintage Books.

Gardner, H. (1987). *The mind's new science*. New York: Basic Books.

Murphy, R. (2002). *Introduction to AI Robotics*. Cambridge, MA: MIT Press.

Shelley, M. (1818; 2003). *Frankenstein*. New York: Penguin Books. (See, particularly, the descriptions of the robotic monster's awakening to the world environment.)

13

Conclusion: Where We Go From Here

"Observe the invincible tendency of the mind to unify. It is a law of our constitution that we should not contemplate things apart without the effort to arrange them in order with known facts and ascribe them to the same law."

—Ralph Waldo Emerson, 1836

The Benefits of Cognitive Science

Cognitive science has made an indelible stamp on the way people now think about mind. Prior to its appearance, there was a plethora of different theoretical approaches to mind. One only needs to look at psychology to see this. Psychology, throughout most of the first half of the twentieth century, generated a profusion of theories on what the mind is and how it should be studied. There was very little common ground that researchers in different theoretical camps could stand on. A psychoanalyst's conception of mental processes was qualitatively different from that of a Gestalt researcher. A behaviorist would have little to say to a structuralist.

The adoption of the cognitive view has meant progress in the area of bringing these diverse theoretical perspectives together. The influence of cognitive

theory can be seen in the ways it has influenced what were formerly considered separate disciplines. There are now cognitive theories of development, social behavior, personality, therapy, and education. Another indication of the integrating character of the cognitive approach can be seen in the ascent of concepts that are implemented across disciplines. The concept of a schema, for example, has equal import in both cognitive psychology and artificial intelligence.

Aside from theory, cognitive science has yielded a multitude of practical applications. Research in artificial intelligence has given us impressive programs that can recognize speech and diagnose complex medical and engineering problems. In the area of robotics, we see the development of new and more sophisticated robots capable of executing complex tasks, ranging from assembly line production to bomb disarmament. These applications have a long-term positive economic impact, as they allow businesses to operate more efficiently and can result in the creation of new industries.

There are practical results that have come out of other cognitive science disciplines as well. Advances in the neurosciences often result in new treatments for disorders such as autism, Parkinson's disease, and Alzheimer's disease. The formulation of cognitive theories in psychology has provided new therapies for the treatment of anxiety and depression. Cognitive insights have also had an impact in education, having led to new methods in the teaching of reading, writing, and other subjects.

Working Memory: An Example of an Integrated Program of Study

In this section we take a single cognitive process and illustrate how it would be studied by each of the different approaches. We do this for several reasons. First, this kind of focus demonstrates how the cognitive science perspective can provide an innovative approach that cannot be achieved by the exercise of separate and independent disciplines. Second, it is a nice illustration of the unique contributions of each approach. Third, it shows what cross-disciplinary links in the study of a specific topic should be like. We do not restrict ourselves to simply describing how each individual approach would study the process, but also explain how each would do so in an integrative manner. The importance of integration is discussed later. The process we choose here is working memory, because it has already been well investigated. It is important to keep in mind that we imply no hierarchy of importance in this exercise; no individual approach has precedence over another.

The Philosophical Approach

A philosophical study of working memory would ask critical questions about the nature of its representations and processes: Is working memory restricted to verbal and spatial representations or can it accommodate others? How are numbers represented? Do any of these representations have meaning? If so, do they have the same sort of meaning as information utilized by other cognitive processes? What are the different functions of the central executive? Can it perform other functions that have not yet been studied? Is there a single central executive used by working memory or are there multiple executives utilized by other processes?

Notice that these sorts of questions need not pertain only to the actual representations and processes that are known to researchers. This is the strength of the philosophical approach. It can address issues that go beyond the delimited boundaries of the phenomenon as specified by the other disciplines. Philosophers are free to ask questions concerning abstract levels of description, beyond even the computational level. These might include metaphysical and ethical inquiries.

The Psychological Approach

How would a psychologist study working memory? We use the term "psychologist" here in the broader sense, referring to researchers and theorists working outside of cognitive psychology. In this case, it would be informative to see what long-standing theoretical traditions have to say on this topic. Using Gestalt theory, we could legitimately ask whether working memory creates Gestalts during problem solving. If so, by what process is a Gestalt formed? Are the Gestalts that are part of visual perception the same as those that are part of the "Aha!" experience that can occur during insight learning?

The Cognitive Approach

Methodology would play a key role in the cognitive approach to working memory. Experiments could be used to test the assumptions of various process models. These models could then be revised on the basis of this feedback and results from other methods, such as computer simulations and brain imaging studies. In Chapter 6 (The Neuroscience Approach) we saw that researchers using PET imaging had discovered new information about the nature of

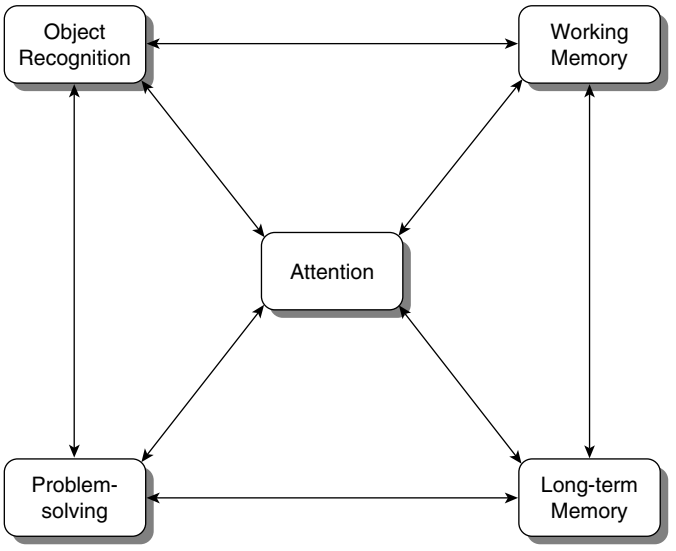


Figure 13.1 Assembling models of different cognitive processes can help us to better understand how each of them operates individually

working memory. This information could then be used to modify existing models.

The cognitive approach lends itself best to explanations at the algorithmic level. Its strength lies in its formulation of detailed process models that show how information is represented and transformed by a cognitive structure devoted to a particular processing task. Witness the large variety of such process models that this field has produced. But these models are for simplicity's sake restricted to a single domain. A new integrative approach would consist of juxtaposing process models and observing how they mutually constrain each other (see Figure 13.1). For example, one could place a model of working memory next to a model of long-term memory and make connections between the two. Matching up the inputs and outputs between them might by itself necessitate structural changes in one or both models.

This technique of model comparison is like putting together the pieces of a jigsaw puzzle. Knowing that the pieces must fit together might require a change in the shape of the pieces themselves. It might also cause us to rethink the nature of modularity. As a result of this sort of exercise, we might find that some processes are quite independent and modular, and that others are more diffuse, with a significant number of interconnections and processing dependencies.

The Neuroscience Approach

The obvious goal for neuroscience in this exercise is to map out completely the anatomy and physiology of the neural systems that underlie working memory. It might seem that this is self-contained knowledge, but as regards comparative neuroscience, a cross-species analysis provides us with a much larger picture. In the neuroscience chapter we saw that the locations for working memory in monkeys and in humans were not the same. The examination of such differences can lead us to infer the functional demands of the environments that have shaped brain structure. A closer examination of the difference between cognitive demands placed on monkeys and those placed on humans, in terms of language, memory load, and so forth, might explain the anatomical differences.

Neuroscience plays another important integrative role—one that we have already mentioned. It places limits on the formation of models designed to simulate some aspect of brain function. Creators of these models would attempt to map the components of the models onto corresponding neural sites. Information processing centers or modules in the models logically map onto nuclei or localized regions of neural tissue where neurons have a high degree of interconnectivity. Connections between modules in turn map well onto tracts of neuronal axons that conduct information from one center to another.

The Network Approach

A network approach to working memory would consist of building network architectures that reproduce various aspects of its function. One could, for example, implement an artificial neural network that reproduces the storage and rehearsal of verbal information. Other networks could simulate the rotation of mental images or the central executive. Larger networks that might integrate these smaller networks into a complete representation of working memory could then be constructed. Although semantic networks are best used in understanding the layout of information in long-term memory, they might also prove useful toward the understanding of some aspects of working memory function.

Like process models, artificial and semantic network models are dynamic. They must undergo modification in response to new discoveries made in other disciplines. The architecture and processing capacities of network models should be altered in response to relevant results from neuroscience and other experimental data.

The Evolutionary Approach

The evolutionary approach puts working memory squarely into the computational level of description. Evolutionary psychologists would ask what purpose different working memory functions serve. They would want to know what adaptive roles these functions play and how they have contributed to the survival of the species over long stretches of time. Evolutionists would of course see working memory as a "bundle" of domain-specific processes or modules, each having evolved to solve a context-specific problem. This way of thinking would stimulate a healthy debate and hopefully an investigation into the issue of whether working memory is a general purpose or domain-specific processor.

The Linguistic Approach

A linguistic perspective on working memory would by definition center on verbal processes. It would examine the role of working memory in language comprehension and production. Because linguistics is such a wide-ranging discipline, it provides multiple perspectives for the study of working memory. A linguist could adopt a developmental perspective, investigating how working memory changes in response to the growth of language ability. In keeping with the linguistic relativity hypothesis, a linguist would investigate if working memory operated differently in speakers of different languages or speakers of more than one language. Linguistic processing in working memory may also be governed or influenced by grammatical rules.

The Artificial Intelligence Approach

Computers already have the equivalent of a working memory. It is RAM, or Random Access Memory, and is used by the machine as a temporary holding buffer for information drawn from a storage or input device. It is no accident that computers and biological organisms must rely on some form of working memory, and this attests to working memory's functional import. An examination of the differences between these systems—those that were designed and those that evolved—and of the different roles they play is needed. It may give us new leads into the computational significance of working memory and may suggest that there are general purpose categories of function.

Artificial intelligence researchers could of course generate algorithms of working memory designed to simulate human or animal capability. They could also design algorithms to solve problems that were apt to be encountered by a

working memory system, for example, that of coming up with the best way to search for a target among the items of a list, or how to match two items against one another. Solutions to these kinds of problems may prove to be optimal from an information processing point of view and may be implemented universally in biological brains.

The Robotics Approach

The role of the physical environment comes into play during the course of robot design. A robotics investigation of working memory would therefore be concerned with how working memory aids in the execution of a complex, real-world task. Working memory in robotics would manifest itself as an interface between perception and action. It could be used, for example, in the service of navigation. Working memory could enable a robot to form a visual representation of a cluttered room and to plot a path that would go from its present position to some goal location. The requirements of a robotic working memory would change with the particular task demands, but in general would subsume an ability to negotiate effectively with the physical environment.

Issues in Cognitive Science

Cognitive science has made new inroads into the understanding of mind. However, there are a number of issues that cognitive science needs to address more completely. One general issue concerns the extent to which cognitive science provides a coherent, unified theory of mind. Also, you may have noticed that cognition is concerned primarily with thoughts as they occur inside a single computing device, biological or machine. Other aspects of mind (for example, emotions and consciousness), the larger context in which mental operations occur, and the significance of physical and social environments all warrant greater attention. In addition, cognitive science gives preference to universal characteristics of information processing systems over individual differences, even though an understanding of the latter is both useful and worthwhile.

A General Issue: Lack of a Single Unified Theory

E. O. Wilson, in his book *Consilience: The Unity of Knowledge* (1999), echoes a sentiment felt by many scientists. He believes that all the sciences

should be unified within a single framework. Cognitive science is a step in this direction. It gives researchers a common set of assumptions. However, cognitive science is not yet a complete, unifying theory of mind. The exact nature of mental representation and computation is still debated. The different approaches still use theories that emphasize different causal and explanatory factors. So, although researchers across disciplines may agree in a general way that the mind is an information processor, they might argue the specifics.

The multidisciplinary approach, while making important advances, has yet to give us definitions of basic terms. For instance, what is mental representation? What is mental computation? The classical information processing view of cognitive science believes representation is symbolic and infers that only select types of transformations on these symbols may occur. The connectionist sees mental representation in a different light. To connectionists, representations are distributed connection strengths in a network and computations result from the activation of nodes and the spreading of activity between them.

The conflict between the classical and connectionist perspectives is a fundamental issue that cognitive science must address. But there are disagreements that crop up when we compare explanations across other disciplines as well. In the evolutionary approach, cognitive processes are explained according to selection pressures acting on multiple generations of a species. In psychology, these same processes can be explained according to developmental, learning, or environmental influences acting within the lifespan of a single individual. So the multidisciplinary approach produces some conflict over the exact nature of mind and how it operates, as well as how to best go about providing explanations for such things. Theoretical diversity, it turns out, is a two-edged sword: what it affords in insights it can take away in overall coherence.

Specific Issues Facing Cognitive Science

Thagard details several challenges to cognitive science. In his book *Mind* (2000) he discusses six problems of cognitive science and possible ways that cognitive science as it exists presently can respond to them. These issues have to do with emotions, consciousness, physical environments, social environments, dynamic systems, and mathematical knowledge. In this section, we spend time exploring some of the questions that Thagard raises. We do not discuss dynamic systems or mathematical knowledge, as they are beyond the scope of this text. We do, however, address the issue of individual and cultural differences because of their close relationship to the role of environments.

Emotions

This book has been concerned with thought processes such as memory, attention, and language. But the mind consists of more than just thoughts. We are capable of experiencing a wide range of emotional states, for example, happiness, sadness, anxiety, anger, and so on. Cognitive science has been censured for having insufficiently accounted for this important dimension of conscious experience. This is actually not the case. Recent years have seen cognitive science beginning to grapple with the topic of emotion. In a recent book, the neurologist Antonio Damasio summarizes a wide variety of case studies (Damasio, 1995); he concludes that emotions are part and parcel of rational thinking and that the absence of emotion can interfere with rationality and can render intelligent decision making impossible. There is extensive ongoing research in the cognitive neuroscience of emotion in which the anatomy and physiology of the structures involved in emotion are being studied in terms of their associations with cognitive processes, for instance, in terms of how emotions are regulated and experienced (Lane & Nadel, 2002).

The evolutionary perspective also offers us an explanatory framework for emotion. According to this view emotions and the capacity for emotion evolved to solve specific problems of survival or reproduction. Fear is a response that motivates us to flee or to avoid dangerous situations (see Figure 13.2). Anger, on the other hand, draws attention to an interfering event and motivates behavior that will contribute to its elimination (Buss, 1989). These two opposite emotional reactions accompany the sympathetic nervous system arousal that occurs during the "fight or flight" response. Similarly, there are theoretical accounts for other emotions. Disgust, for instance, has been interpreted as an evolved mechanism that promotes the avoidance of substances bearing disease-causing agents (Rozin & Fallon, 1987). If these accounts are correct, emotional mechanisms evolved and exist alongside cognitive ones. Just as there seem to be cognitive modules for language or problem solving, there may be emotion modules that mediate action in a specific context. The task for cognitive science, then, is to incorporate these emotion modules into models of mind and to specify how the new models interact with the cognitive models already postulated.

One problem with this formulation of new models concerns the qualitative differences between cognitions and emotions. Thoughts seem to be subjectively "neutral" and lend themselves well to symbolic representation and computation. Emotions, though, seem to be a completely different breed of animal. The feeling of what it is like to be angry or sad seems far removed from the conscious experience of a thought. This means that emotions may not be best represented in the ways thoughts are—in the form of propositions, images,

Figure 13.2 Fear is an adaptive emotional response. It motivates us to avoid dangerous situations. Emotions like fear interact with and influence our cognitions

analogies, and so on. This poses a challenge for model making, because, as we saw in our discussion of analog versus digital representations, the form of the representation to some extent determines the way it is processed. Assuming that emotions do require a different kind of representation, they in all likelihood require a different set of computational rules.

Emotions differ from thoughts in several other important respects. Emotional responses involve the body and not just the brain. These physical changes include alterations in body temperature, hormonal secretion, heart rate, and muscle tension (Levenson et al., 1992; Witvliet & Vrana, 1995). Any account of emotions must therefore take into consideration the wider physiological changes that occur outside the brain. This could be achieved by introducing models that allow for bodily feedback during the processing of emotion.

Thagard (2000) also brings up the issue of neurotransmitters in regard to emotions. Many emotions rely on particular neurotransmitters. Low serotonin levels, for example, are implicated in clinical depression and feelings of sadness. The administration of the SSRI drugs (Selective Serotonin Reuptake Inhibitors) or other antidepressant medications boosts serotonin activity, producing an alleviation of symptoms. It appears that there are other neurotransmitters that play specific roles in the expression of or experience of particular emotions. Current artificial neural network models assume that all neurons act in the same way. These simplified assumptions need to be discarded if we are to take into account the differential effects of separate neurotransmitter systems. This could be implemented in an artificial neural network model by giving different operating characteristics to subsets of nodes corresponding to different transmitters.

Consciousness

In Chapter 2 (The Philosophical Approach) we defined consciousness, roughly, as the subjective experience of our mental states. Cognitive science, with its emphasis on representation and computation, is good at describing the kinds of structures and processes that underlie thought. But it is now turning its attention to investigating what it is like to be in possession of those structures and processes. As we saw in Chapter 2, there are now a number of theories of consciousness. There has also been work on the topic that has emerged from a variety of perspectives, ranging from evolution to cognitive psychology to neuroscience (Blackmore, 2004). The subjective phenomenon of consciousness, it turns out, is amenable to being studied scientifically and objectively.

One way of explaining consciousness is to equate it with attention. In this view conscious experience is where attentional resources are allocated, whereas subconscious or unconscious processing happens without attention. Johnson-Laird (1983) proposes that consciousness is like the operating system of a computer that diverts processing resources to the most important tasks. In human (and other) minds consciousness may play a similar role, shifting attention to what needs to get done right now. One can see the need for such a central executive, given the brain's highly parallel and divided processing nature.

The idea of consciousness as a central executive that coordinates other mental activities already exists in several cognitive models. Think back to the executive in Baddeley's model of working memory. Specifying the role of a central processor is one way of supplementing other cognitive models. The introduction of the executive could take the form of adding in an executive function

control box that could turn other functions on or off and regulate their activity in various ways.

An alternate way of dealing with consciousness is to equate it with brain activity. As discussed in Chapter 2, Crick and Koch (1995) posit the existence of specialized consciousness neurons. We also saw in that chapter another neuroscience model for conscious experience that was based on a thalamo-cortical circuit (Churchland, 1995). In that model, consciousness is the activation of a neural circuit. This of course is a monistic, materialist, and functionalist answer, since it says that conscious experience, like any other thought process, boils down to brain activity. This neuroscience view requires us to adopt a philosophical stance that, in the opinion of some, is not a solution because it defaults to one side of the debate.

Physical Environments

An important comment on cognitive science is that minds, unlike computers, exist in the context of a complex physical world. In this conception, mental activity does not occur in a vacuum, isolated from the surrounding world. Much of our thinking is directly connected to sensory inputs and motor outputs. It is devoted to interacting with the "outside," as opposed to operating only on complex forms of representation and computation generated from the "inside." This idea is known as **embodiment**. Figure 13.3 shows cognition as one aspect of a world-interaction process. External stimuli that are interpreted by perception can lead to action directly (as when we reflexively catch a ball that is coming at us) or indirectly through cognition (as when we deliberately make a decision to throw a ball). The actions, in turn, alter our perceptions, which further alter our actions (once the ball is thrown we can track it to see where it has gone). This idea is in contrast to a static view of the senses, according to which they just passively take in information. The action-perception cycle of Figure 12.5 in Chapter 12 (Robotics) depicts a similar idea.

The physical environment argument, for some tasks at least, seems to do away with the idea of representation. Dreyfus (1992) contends that intelligence, and consequently the performance of efficacious action, do not require formal symbolic representation. Much of what we do can happen as the result of an interaction between an agent and the world. For instance, imagine picking up a pen from a desk. This need not require a visual image or linguistic representation of the pen in memory. The perception of the pen generated from the sensory systems is enough to guide an arm toward the pen to grasp and lift. Simple learning principles based on such sensory-motor experiences are enough to allow for sophisticated interaction with a complex environment. We have seen this idea already. Brooks's (1991) Subsumption Architecture, discussed in

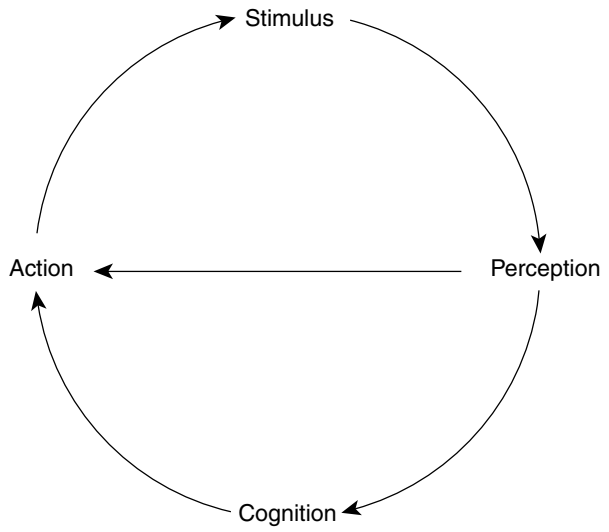


Figure 13.3 A cyclical perception-action model of the perceptual process

Chapter 12, is based on the idea of learning through interaction. The robots constructed with this architecture have been quite successful in locomoting about the world.

Perception is a good example of where the representational-computational and sensory-motor views contrast. The dominant view of perception is that it is computational. All the theories that were summarized in Chapter 4 (The Cognitive Approach I) involved a complex series of stages where features were extracted and used to reconstruct an internal representation of an object. This process is inferential and indirect because the cognitive system must recognize or infer some property of the object through a time-consuming set of processes.

James Gibson (1986) proposes instead that perception is direct and immediate. He argues that perception results from the interplay of an organism with the world. In his view, the visual system uses information in the image directly in its carrying out of perceptual tasks, without resort to representations and computations. For instance, our ability to judge our position in space while moving comes from optic flow, whereby surfaces move in the visual field in a coherent fashion (see Figure 13.4). Our ability to judge the sizes of objects relies on where they are cut off by the horizon line, which in effect equals our eye height. Properties such as optic flow and eye height are examples of information that is directly available to us. They do not require inferences or extended computation. As such they can be considered the perceptual analog

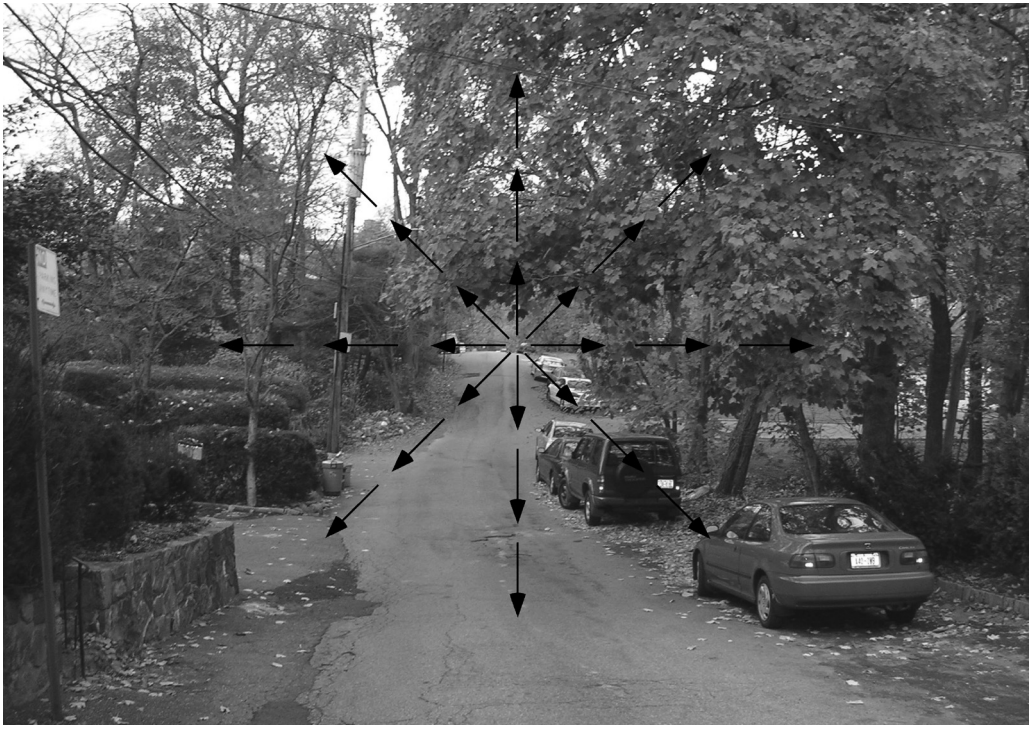


Figure 13.4 Forward locomotion produces a global optical expansion of texture in the visual field, indicated here by these arrows. Information about speed is provided by the rate of expansion

of heuristics in problem solving. They provide quick and, in this case, accurate solutions to perceptual "problems." Gibson's work established the field of **ecological perception**. This approach provides an alternate framework for describing perceptual processes.

Social Environments

There is a famous quote that goes: "No man is an Island, entire of itself; every man is a piece of the Continent, a part of the main" (from a poem by John Donne written in 1624). By this Donne meant that we live in a social environment in which we interact with family members, friends, coworkers, and others on a regular basis. A social environment from a cognitive point of view is one where individual minds exchange information.

In these sorts of environments cognition takes place to some extent outside the individual. Novel information or the solution to a problem often arises from group interaction. We see this in the business world, where a new marketing campaign or a sales strategy results from a company team working collaboratively on a project. The concept of thought emerging as a result of information exchange between individuals is known as **distributed cognition.** If you can imagine the boxes in Figure 7.1 (from Chapter 7) replaced by individual minds, you will have a general idea of what a distributed cognitive architecture is like.

The computer analog to social cognition is **distributed artificial intelligence.** Here, individual computers with separate processing capabilities are linked and communicate with one another in a cooperative fashion. If the individual computers have different specializations or computing strategies, they can usually solve problems together that none of them could solve alone. Similarly a group of individual persons, each having his or her own knowledge and ways of thinking, is more likely to arrive at a solution than any one of those persons in isolation.

These examples from the human and computer domains show us the power of cooperative computing. But what constitutes the "individual" unit in an information processing system? Marvin Minsky, cofounder of the MIT AI laboratory, proposes that the mind is itself the product of countless smaller processing units—what he calls **agents**. Each agent performs a very limited and specific task, but working together, they give rise to more complex cognitive capacities. In his book *The Society of Mind* (1985), Minsky describes how cooperative agents can produce many cognitive phenomena, including pattern recognition, memory functions, and language. See the In Depth section for more on his theory and its relation to multiagent systems.

The concept of a society of mind is echoed in the parallel processing architecture of the brain and in Dennett's (1991) Multiple Drafts model of consciousness. It forces us to rethink what we mean by mind. Is mind a single unitary entity or just a collection of smaller computing units? It also forces us to consider the hierarchical structure of mind, wherein computational elements are nested one inside the other. Going downward toward the microscopic, mind is made of neural circuits and neurons. Going upward toward the macroscopic, we can now see that it is made of physical and social environments.

A way of folding the social environments challenge into the cognitive science view is through the use of **shared representations**. These are representations that can be processed by more than one user. Language is the medium by which shared representations exist. The written and spoken forms of language allow people to hold the same sets of ideas in common. Language converts

representations that may exist only in one person's mind into a form that can be transmitted to and comprehended by others.

We can now begin to elaborate on what models of a social cognitive science would be like. They would have to indicate the individuals in a group and the representations and computations each could perform. This much is already in place. Additionally, they would have to formalize the representations shared by individuals and how they are processed differently. They would also have to indicate how the results of an individual computation could produce new shared representations. This is especially important, because it is the differential processing and subsequent re-communication of shared representations that give rise to emergent ideas.

Individual and Cultural Differences

Two other issues, related to the role of environments, are individual and cultural differences. One of the first and most obvious conclusions that we draw after interacting with others is that they are different from us. Individuals have unique and in some cases radically different ways of thinking. This begs many interesting questions: How is it that people come to think differently? Is there a better or worse way of thinking? What is it that makes an acknowledged genius in a field different from the average person?

This phenomenon on a larger level comes into view when we start to compare societies, cultures, or countries. Of course, differences exist here as well. Segall, Campbell & Herskovits (1966) found that Americans and Europeans were more prone than other world populations to experience several visual illusions, and they attributed this difference to the exposure of Europeans and Americans to "carpentered" environments that contain many rectilinear forms. There is some evidence to suggest that cultural experience plays a role in the perception of space (Deregowski, 1989). There are also fundamental personality differences among those living in different cultures. The Japanese and Chinese, in comparison to people from western cultures, are shyer toward strangers and exhibit greater concern for social harmony and loyalty (Bond, 1988; Triandis, 1994). These studies suggest that culture plays an important role in shaping perception and cognition.

The historical trend in cognitive science has been the study of universal characteristics—what all people or cognitive devices share in common. Research efforts have traditionally focused on solving problems that every information processor in a particular domain must solve, such as the object constancy problem in vision or the consolidation problem in memory. Cognitive science has devoted less attention to the way these processes may be carried out in a lone individual.

One way to study cultural differences is **anthropology**. Anthropologists investigate when and where humans first appeared on the earth and why they have varied with respect to some physical traits. They are also interested in how and why societies vary in their customary ideas and practices (Ember & Ember, 1985). Because of its focus anthropology is a useful adjunct to the evolutionary perspective. This discipline can also tell us a lot about which thought processes are universal and which are particular to specific cultures. Anthropology can help us to elucidate the factors that account for such differences. Anthropologists who have made significant contributions to our understanding of mind include Lucien Levy-Bruhl, Franz Boas, Claude Levi-Strauss, and Dan Sperber (Gardner, 1985).

The quest to understand universal phenomena is certainly important and is arguably the first goal of an emerging science. A prime motivation for this quest is that it helps us understand the basic processes involved. The details can be worked out afterward. However, it is sometimes these very same details that can illuminate the operation of the more basic processes. For example, understanding why some people are so much better at math than others can probably tell us a lot about basic mathematical ability and the function of working memory (see Figure 13.5). Research on universals and particulars, therefore, ought to proceed in parallel. In this fashion each can inform the other.

Enhancing Cognitive Science

In the previous section, we discussed some of the general and some of the more specific problems facing cognitive science. Included in this discussion were suggestions for how the field as a whole might benefit. For instance, we decided that if emotions are a fundamental aspect of mind and influence cognition, then cognitive science ought to take them into account by generating new forms of representation and computation to accommodate them. In this section we review a number of additional suggestions for how cognitive science might be enhanced. We first discuss theory and outline five criteria for evaluating theory. We then move to a discussion of integration of various kinds: across levels of explanation, disciplines, and methodologies.

Evaluating Cognitive Science Theories

An important issue in cognitive science centers on how we go about evaluating different theories. Since cognitive science is such a far-ranging endeavor that spans multiple approaches, the number of theories that are generated is

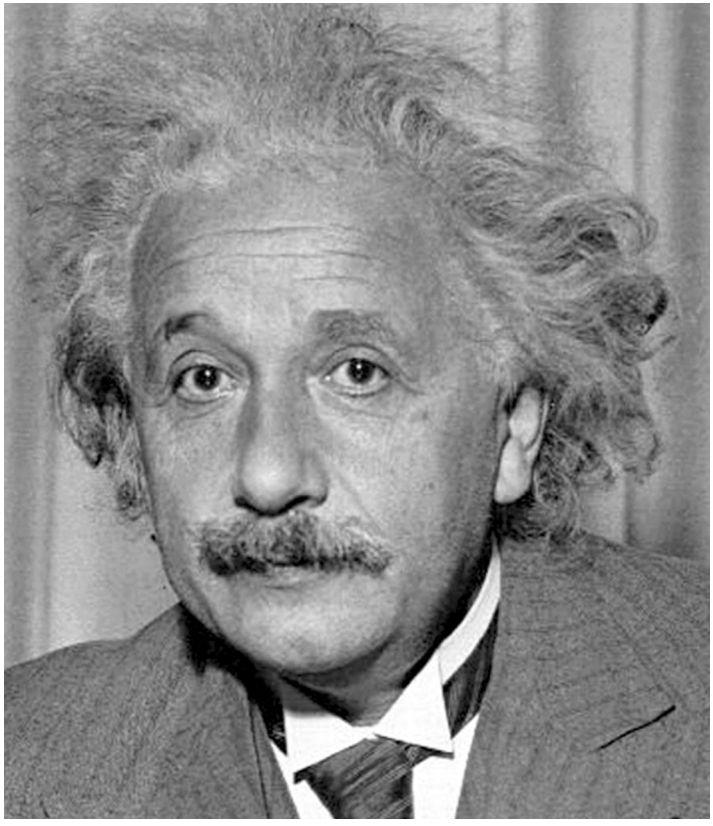


Figure 13.5 The physicist Albert Einstein. Isn't it worth understanding how his mind differs from that of the ordinary person?

quite large. The field could therefore benefit from an efficient method for determining what is a good or bad theory, as well as what is an unacceptable theory. Thagard (2000) proposes five criteria by which we can evaluate theories of cognitive representation. These criteria are equally effective guidelines for judging theories of mind. Let us define and discuss each of these briefly. Notice that there is one discipline that is best suited to the use of each criterion.

1. *Representational power.* How well does the theory account for the wide variety of cognitive representations? As we have seen, representations come in

many forms. Images are good at representing spatial layout and lend themselves naturally to certain types of processing, such as navigational computation. But images have a hard time representing abstract concepts such as "democracy." These concepts are best captured using linguistic forms of representation that are inherently superior for other sorts of computation such as logical reasoning and inference. Philosophy can give us good insights into the nature of representations and their role in different types of processing. A good cognitive science theory is one that can account for different types of representations.

2. *Computational power*. How well can the processes specified by the theory under consideration perform a given computation? For any given cognitive problem or task, there may be several ways to achieve a solution. The effectiveness of a given solution can be evaluated on the basis of its speed and flexibility. In general, speedier algorithms that can complete their operations faster are better. So are those that can solve a greater variety of problems and are not just restricted to one domain. Computer science and artificial intelligence are fields that excel in the evaluation and testing of different algorithmic solutions. We have to be careful at this point to distinguish between practicability and plausibility. A fast and flexible algorithm may be of great use in software application, but have no bearing on how the processes the software executes are performed in humans or animals. For this, we must turn to considerations of plausibility.

3. *Psychological plausibility*. How likely is it that the processes specified by the theory are actually carried out? This question can be answered by rigorously weighing theoretical assumptions against empirical evidence. Experimental psychology is good for this. Data from experiments can rule out incorrect mental models and suggest which types of mental computation are being performed. We cannot, however, rely solely on experimental psychology to test plausibility. This is because the results from these experiments can only validate theoretical constructs at an abstract, information processing level of description—Marr's computational and algorithmic levels. To get an implementational or hardware level of description we need to take biology into account.

4. *Neurological plausibility*. How are the mental processes physically carried out in a neural substrate? An elegant theory that can explain all of the psychological data is hollow if it cannot be implemented biologically. The neuroscience and network approaches fill this gap. Brain scanning studies tell us which overall parts of the brain are involved in performing a cognitive task. Network models give us a more local account of how information may be represented and computed.

5. *Practical applicability*. How can the knowledge gained from a cognitive science theory be put to use in the real world? This issue reflects the longstanding difference between theoretical and applied science. Theory is important for understanding natural phenomena. The knowledge gained from theory can then be used to solve social problems in engineering, education, medicine, and other areas. A good theory is one that will yield good practical applications.

The Role of Integration

We have seen, throughout this book, that each cognitive science approach, by itself, yields a substantial quantity of information about mind. We have also seen that an integrative approach, in which the disciplines work collaboratively together on a common problem, is an even more powerful tool for investigating phenomena of mind. The working memory example described above demonstrates just this. But this form of integrative research isn't easy. This is because the different disciplines vary with respect to explanatory constructs and methodology. In this section, we propose kinds of integration from which cognitive science research could benefit. These are integrations across levels of description, disciplines, and methodologies.

Integration Across Levels of Description

Recall again David Marr's three levels of description that make up the tri-level hypothesis. These are the implementation level, corresponding to a hardware explanation of a cognitive process; the algorithmic level, at which how information is processed is specified; and the computational level, at which a definition of the problem and its solution are given. These levels, in the order given, correspond roughly to three basic questions about a cognitive process: Where is it taking place? How is it taking place? Why is it taking place?

It is rare in cognitive science to find a theory that provides an explanation at all three levels. Rather, different approaches tend to emphasize one level over another. Neuroscience and robotics focus on the implementational level, since they either study the brain's existing architecture or design new mechanical and electronic architectures. The cognitive and artificial intelligence approaches, because of their models and programs specifying information flow, emphasize the algorithmic level. The philosophy and evolutionary approaches focus on the computational level, since they provide richer explanations for the purposes of cognitive processes.

What is needed, then, are more theories or models that integrate across all three levels. The levels would then mutually constrain one another, leading to

better and more accurate formulations. For instance, in the neuroscience approach, we saw that the Treves and Rolls (1994) model of hippocampal function was based on what we know of the anatomy of the region. This is an instance of integration across the implementation and algorithmic levels. Evolutionary accounts of hippocampal function could further enhance their model. Knowing why the hippocampus evolved and under what specific contextual circumstances it operates would provide further constraints.

Integration Across Disciplines

It is difficult to formulate a cognitive science theory that satisfies all five of the theoretical constraints discussed earlier, and few researchers actually try. One reason is that most investigators are specialists. They are trained in the knowledge and tools of their own discipline, but lack understanding of other allied disciplines. An artificial intelligence researcher may be adept at devising search algorithms to locate information in a database, but not be familiar with the literature on human memory search (an area of expertise reserved for a cognitive psychologist). Similarly, a neuroscientist may know which brain areas underlie memory retrieval, but not be able to implement a neural network model of the process (a skill possessed by someone working within the network approach).

There are two solutions to this. The first is to train individual researchers in the interdisciplinary perspective. Scientists who possess an understanding of cognitive science fields outside their own can then formulate more appropriate theories. This solution has its limitations, however. Scientific practice requires a high degree of specificity and focus, and researchers cannot be expected to acquire more than rudimentary training in multiple alternative disciplines. The most that can be expected is some degree of familiarity with the literature and techniques of disciplines outside one's own specific area.

A second solution (mentioned in the introductory chapter) is interdisciplinary cooperation. Here, investigators from different disciplines work together on a common project. For instance, a team made up of investigators from different disciplines might focus its research efforts on a single topic, such as the motor action of grasping. Each investigator can then contribute the unique strengths of his or her discipline's theoretical insights and methodology. The intercommunication that results can lead researchers to ideas they might not have had otherwise. Recently formed interdisciplinary cognitive science research centers sometimes employ philosophers, cognitive psychologists, computer scientists, and neuroscientists. Studies at these centers have yielded findings unlikely to have been obtained by departments operating autonomously.

Integration Across Methodologies

Many of the disciplines employ distinct methods. Philosophy uses logical reasoning; psychology and others, the scientific method; cognitive psychology, artificial intelligence, and the network approach use modeling; neuroscience utilizes a combination of case studies, the lesion method, and brain imaging, whereas robotics builds and tests machines in a real-world environment. Each of these techniques has its strengths and weaknesses. A computer simulation of a cognitive process may yield solutions to the problem of how to best represent information and of how to solve a problem in the smallest number of processing steps. This simulation, however, would ignore empirical data at its own risk, since nature may have evolved a different or better solution.

In Chapter 4 we discussed several models of attentional selection—some favoring early selection, others favoring late selection. Different models were formulated over time, with models being modified to accommodate new experimental data. This is a good example of integration across two methodologies—modeling and experimentation. Although in this case the models were modified in response to experimental results, the relationship is really two-way, with models providing the impetus for designing new experiments as well.

The Future

A few parting words. Cognitive science is unique in the history of science in that it puts under one umbrella such a large number of disciplines. This provides an unparalleled opportunity for these disciplines to work together in new ways. Cooperation and competition among the various approaches are both necessary to the advancement of our understanding of mind.

But what does the future hold in store for cognitive science? Where are we going from here? Progress in any area of science relies on a number of factors. Among these are theoretical breakthroughs that reorganize our understanding of facts as well as technological advances. Cognitive science, perhaps more than some other disciplines, is dependent on new tools for measuring or simulating mental processes. Recent years have seen the development of such tools, including artificial neural networks and fMRI imaging techniques. It is likely that other new tools like these are on the way. It is not unrealistic to imagine brain-imaging devices that will be able to record the individual neuronal activity of large cell populations or instruments that will be able to measure quantum fluctuations at the atomic level. The use of such tools may open up a whole new vista of mental territory rife for exploration. The future in cognitive science is bright indeed.

In Depth: Multiagent Systems

An agent is a component module of a computer system that is endowed with the ability to perform some task. Collections of different agents interacting with each other form a **multiagent system** (Wellman, 1999). Although individual agents considered in isolation follow fairly simple rules, the system as a whole is capable of sophisticated information processing activity and may be considered intelligent. Distributed artificial intelligence systems are in fact a type of multiagent system (Bond & Gasser, 1988).

Agents interact with each other based on an interaction mechanism that specifies the kinds of information they can exchange and their method of doing so. In some cases the mechanism allows for cooperative agent interaction, meaning that the agents all have the same goals, although they may differ in the roles they play as they try to achieve them. Agents in a cooperative system negotiate with one another: they exchange information in order to accomplish their tasks.

There are also noncooperative systems in which agents have different goals. In these systems the agents still cooperate, but only when it is in their own "best interest." That is, an individual agent will negotiate with another only when such negotiation furthers its own unique objectives. Agents in noncooperative systems are usually programmed to act in accordance with principles of rational decision-making. They choose from among a set of possible alternatives, ordered by preference. Economists study this type of decision-making process in order to understand phenomena such as consumer choice (Henderson & Quandt, 1980).

The hallmark of a multiagent system is decentralized decision making. Groups of agents at one level or location in the system can make a decision or solve a problem using partial information. This is in contrast to a centralized system, in which a central executive performs these duties after having all available information fed to it. A decentralized organization makes multiagent systems more robust and more resistant to damage than centralized systems. For this reason, they operate effectively in network applications, such as telephone and computer systems. In these systems, if part of the network goes down, agents in the remaining portion can determine how to reroute signals.

The most obvious metaphor for a multiagent system is a human society in which the agents are individual persons who make decisions and then act on them as a result of communication with others. Individuals can of course be cooperative or noncooperative and may have shared or individual goals. Human societies can also be centralized, as in a military dictatorship, or decentralized, as in a democracy. The emergent behavior of a multiagent system, as we have seen, can solve problems and act as if it were intelligent. Similarly, the

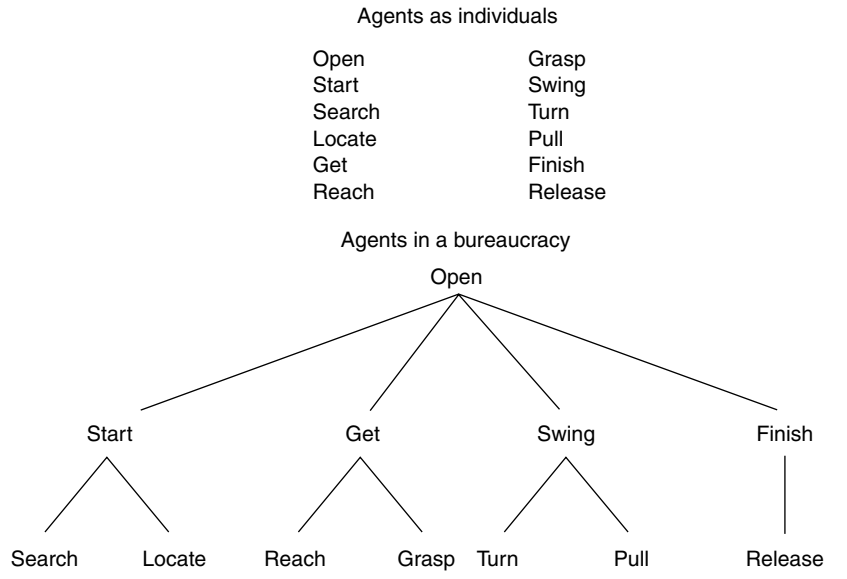


Figure 13.6 Agents by themselves and in a hierarchical relationship that opens a door

cumulative actions of a human society can also be conceptualized as solving "problems" of law, commercial exchange, and military conflict.

Multiagent systems are a theory of mind, because the mind, like a society, may be considered as having interactive agents (Minsky, 1985). But what are these mental agents? The modules we have discussed throughout this book are one candidate, although, because of their functional independence, they are limited in the extent of their interaction. Other agents could be neural circuits that control a specific function. This is what Marvin Minsky suggests agents are. He lists four agents that would allow us to pick up a cup of tea: a "thirst" agent that makes us want to drink the tea, a "move" agent that wants to get the cup to our lips, a "grasp" agent that enables us to maintain our hold on the cup, and a "balance" agent that wants to keep the tea from spilling out.

Notice that the agents in the preceding example reside in different systems—some for motivation, others for motor action. These agents also call on each other in a variety of ways, both linear and hierarchical. The "move" agent could call on another "reach" agent that would then call on the "grasp" agent. The "grasp" agent would then call on lower-level agents that move the fingers, and so on. Figure 13.6 shows the relationship between agents in a hierarchical organization. This example shows that the complexity of behavior that emerges from multiagent systems comes not from the agents themselves, which

have very little actual computational power. It comes from the way they are connected to and control one another. The distributed processing nature and high degree of connectivity in the human brain certainly give it the potential to act as a multiagent system.

Minds On Exercise: Evaluating Theories of Mind

Pick an important cognitive process such as visual object recognition and describe how it would be studied from each of the approaches in this book. Do some of the approaches seem to lend themselves well to the study of the process? Why? Are there some approaches that make it difficult to study the topic? Why? Describe what a cross-disciplinary study of the process would be like. You can do this by picking any two approaches at random and detailing what a collaborative effort between them would be like.

Food for Thought: Discussion Questions

1. What is an emotion? Would you make a decision differently if you were angry (versus being sad)? How would it be different?
2. Does cognitive science need to concern itself with the issue of consciousness? Is it enough to explain mental processes objectively, or do we need to account for their subjective character as well? Would you be satisfied with a science of mind that could explain every mental event perfectly but not what it is like to experience them?
3. Do you think it would be possible for a mind to develop without interacting with a physical or social environment? Could a computer ever become conscious or intelligent via direct programming from its creators and in the absence of being allowed to learn, compete, and grow in response to the demands of an environment?
4. Take any one of the theories of mind from this text and evaluate it using Thagard's five criteria. In what areas is it strong? In what respects is it weak? Modify the theory to improve it.
5. In what way is a city like a multiagent system? List some of the possible agents in a city and how they might interact with each other. Organize the agents into different groups based on their functions. Which agents or groups interact with each other? Which ones do not? What kinds of "problems" must a city solve?

CHAPTER REVIEW AND EXTENSIONS

Log on to the student study site at http://www.sagepub.com/csstudy for electronic flashcards, review quizzes, and a list of Web resources to aid you in further exploring the field of cognitive science.

Suggested Reading

Blackmore, S. (2004). *Consciousness: An introduction*. Oxford, UK: Oxford University Press.

Dalgleish, T., & Power, M. J. (Eds.). (1999). *Handbook of cognition and emotion*. Hoboken, NJ: Wiley.

Damasio, A. R. (1995). *Descartes' error: Emotion, reason, and the human brain*. New York: Quill.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Lawrence Erlbaum.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Harvard University Press

Minsky, M. (1985). *The society of mind*. New York: Simon & Shuster.

Weiss, G. (2000). *Multiagent systems: A modern approach to distributed artificial intelligence*. Cambridge, MA: MIT Press.

Wilson, E. O. (1999). *Consilience: The unity of knowledge*. New York: Random House.

Glossary

Acoustic code A memory code based on the sound of the items.

Act A robotic primitive that pertains to the actuating or motor elements of an intelligent agent.

Action potential The electrical signal that results when the inputs to a neuron exceed its threshold of excitation.

Action schemas In the Norman and Shallice model of attention, units that govern the execution of context-specific actions.

Activation function A function that maps the strength of the inputs a node receives onto its output.

Adaptive Resonance Theory (ART) network An unsupervised multilayer recurrent network able to classify input patterns into different categories without a teacher. It does this by employing a resonant dynamic, in which the patterns in the first and second layers repeatedly stimulate each other until the pattern in each layer settles into a stable state.

Affordance Perceivable environmental elements that are suitable triggers for an action.

Agent link A link in a propositional network that specifies the subject of the sentence, the one performing some action.

Agents A component module of a cognitive system that is endowed with the ability to perform some specific task.

Agraphia A deficit in writing.

Alexia A deficit in reading.

Algorithm A formal procedure or system that acts on informational representations.

Analog representations A form of representation in which information is coded in a continuous way.

Analogy A mental representation used to compare known information with new information to which it is similar.

Android A robot or mechanical device that resembles a human being in all respects, including the cognitive attributes.

Angular gyrus A brain area located posterior to Wernicke's area. Damage to this area can produce problems in reading and writing.

Anomia A form of aphasia in which patients can recognize and define words but are unable to name them.

Anterior A direction indicating toward the front.

Anterograde amnesia An inability to remember new information after a traumatic incident.

Anthropology The discipline that studies cultural differences by investigating when and where humans first appeared on the earth and why they have varied in some physical traits. Anthropology is also interested in how and why societies vary in their customary ideas and practices.

Aphasia A neurological language deficit.

Apperceptive agnosia A type of agnosia in which people cannot assemble the pieces or features of an object into a meaningful whole.

Appropriate causal relation The relation between sensory inputs, motor outputs, and the representations that come between them.

Arbitrary The property that there need be no relationship between linguistic symbols and their referents. Any symbol could be used to stand for any concept.

Architecture Functional organization of an entity, possibly a computer.

Arcs Represent relationships between concepts or elements.

Arcuate fasciculus A pathway that connects Broca's and Wernicke's areas.

Arousal The extent of physiological activation reflected in measures such as cardiac and respiratory rate.

Articulatory loop A system in the working memory model where speech and sound-related information are rehearsed.

Artificial life The study of manmade systems that behave in ways characteristic of natural living systems.

Artificial neural network (ANN) A computer simulation of how populations of real neurons might perform some task.

Assertives A type of speech in which the speaker asserts his or her belief.

Associative agnosia A type of agnosia in which people perceive a whole object but have difficulty assigning a name or label to it.

Attention Concentrated mental activity in which processing resources are allocated to information sources.

Automatic attentional processes Processes that do not require conscious attentional control.

Autonomous Being able to function without recourse to a human operator.

Axon A long, tubular structure that projects outward from a neuron cell body and serves to transmit the action potential over distance.

Babbling stage A stage in language development where infants produce many sounds occurring in the languages of the world.

Basal ganglia A collection of brain areas involved in voluntary motor responses.

Base-rate fallacy A neglect of the probability that an item belongs to a particular category.

Basis function A function specifying the amount of stimulation a given node receives. It is determined by multiplying each weight by the amount of activation it receives and then summating.

Behavior A mapping or translation of sensory inputs into a pattern of motor actions intended to complete a task.

Behavior-based approach A type of problem-solving method in which a network is allowed to produce a solution on its own. This does not involve the use of symbols. It is representative of the connectionist approach in cognitive science.

Binding problem The issue of how to recombine the various features of an object that have been processed by distributed brain areas.

Biological plausibility The idea that artificial neural networks effectively represent and model characteristics of real-world brains.

Bottleneck theories Theories that describe why it is that of all the information presented to us, only a small amount actually gets through to conscious awareness.

Brain imaging A technique that allows researchers to see the static three-dimensional organization of brain areas and the dynamic activity of these areas over time.

Breadth-first search A methodology for exploring a knowledge base in which one level of the information is completely exhausted before proceeding to the next, more-specialized level.

Broca's aphasia A language deficit in which comprehension is intact but there are impairments in speech pronunciation and production.

Broca's area A region in the left frontal lobe that underlies language production. Damage to this area results in Broca's aphasia.

Bus Informal designation of a computer's communication system.

Capacity theories Theories that conceptualize attention as a limited resource that must be spread around to different informational sources.

Cartographer A software structure that contains all the information needed for robotic navigation—the Knowledge Base of the Hierarchical Model.

Case (Record) An instance of an activity or event.

Case study A method in neuroscience in which researchers examine the relationship between brain damage and behavior in individuals.

Catastrophic interference This occurs when a network has learned to recognize an old set of patterns and then is called on to learn a new set. The learning of the new set modifies the weights of the network in such a way that the old set is forgotten.

Cell assembly According to Hebb, a small group of neurons that repeatedly stimulate themselves.

Central processing unit (CPU) That part of a computer that executes the instructions and controls the sequence of instructions to be carried out.

Cerebral hemispheres The two halves of the cortex, each with corresponding functional specializations. The left hemisphere is more analytic, serial, and logical. The right hemisphere is more synthetic, parallel, and relational.

Cheater detection The ability to detect who has undeservedly received a benefit.

Chinese room scenario A hypothetical situation in which a man uses a set of instructions to produce replies to questions in Chinese. It argues that the man can never learn to understand Chinese.

Chunking The grouping of items together into a single meaningful whole in short-term memory.

Cingulate cortex A region of the cortex that is implicated in selecting a response, especially when that response requires inhibiting or ignoring another alternative response.

Class membership The extent to which, or belief that, an entity belongs to a given class.

Classical conditioning A form of learning in which two repeatedly paired stimuli become associated so that one that formerly did not elicit a response now does.

Classical dualism The belief that both mental and physical realms are possible where the mind controls the body.

Clearness The clarity of a sensation determined by the amount of attention paid to it.

Closed world paradigm References environments in which a model of the world is self-contained.

Closure The perceptual principle stating shapes forming closed objects go together.

Coarticulation Differential pronunciation of a given phoneme based on the phonemes that come before or after it.

Cocktail party effect The ability to automatically hear one's name spoken at a distance even though one's attention is focused on the immediate surroundings.

Cognitive architectures Specify the structure and function of many different cognitive systems and how they interact.

Cognitive economy The principle that concepts should not have to be coded for more times than is necessary.

Cognitive map A mental representation that can explain learning without resorting to pure stimulus-response associations.

Cognitive neuroscience The study of the physiological structures and processes underlying cognitive function.

Cognitive penetration When one's knowledge, beliefs, goals, or other cognitive states alter performance of a mental task.

Cognitive psychology The study of human knowledge representation and use.

Cognitive science The scientific interdisciplinary study of the mind.

Commissives A type of speech that commits the speaker to some later action.

Communication The production, transmission, and comprehension of information between individuals in a language.

Compatibilism The belief that free will and determinism can be reconciled or made compatible with each other.

Complex ideas Ideas formed from the active mental combination of simple ideas.

Compositional semantics When the entire meaning of a sentence is derived from the meaning of its parts.

Computation The transformation or manipulation of information.

Computational level An abstract level of analysis that asks what type of problem a computation solves and how it may have arisen.

Computer axial tomography (CAT) A brain imaging technique in which beams of x-rays pass through the skull and are picked up by a detector on the opposite side. By collecting the information from each of these beams, two- and three-dimensional views of the underlying tissues can be reconstructed.

Concept An idea that represents a class of entities that have been grouped together.

Conceptually driven process A process where context and higher-level knowledge aid in recognition.

Conditioned response (CR) In classical conditioning, a response elicited by a conditioned stimulus.

Conditioned stimulus (CS) In classical conditioning, a stimulus that elicits a response only after being repeatedly paired with an unconditioned stimulus.

Conduction aphasia Results from damage to the arcuate fasciculus. Characterized by a difficulty repeating words that have just been heard.

Conjunction fallacy A neglect of the conjunction rule, which states that the probability of simultaneously being a member of two categories is always less than the probability of being a member of either category alone.

Connectionism The study of mental operations through the construction and testing of artificial neural networks.

Conscious That aspect of mind that contains those thoughts and feelings of which we are aware and can directly access.

Consciousness The subjective quality of experience.

Consolidation The process by which information is transferred from working or short-term memory to long-term memory.

Contention scheduling The scheduling that governs routine habitual performances.

Contralateral Refers to the opposite side of the brain or neural structure.

Control group The group in an experiment that does not receive the independent variable.

Controlled attentional processes Processes that require conscious attentional control.

Convergent dynamics The state of a network represented by how the values of its weights change over time.

Cooing stage The earliest period in language development when infants exercise their vocal cords and mouths prior to articulation.

Coronal plane A cut through the brain that separates it into a front and back piece.

Corpus callosum A collection of fibers that connects the two cerebral hemispheres.

Creative synthesis A principle by which the mind actively organizes mental elements together such that the resulting whole contains new properties. These new properties cannot be explained by the characteristics of the individual elements themselves.

Critical period A period of development during which linguistic experience is crucial for future language use. If a child is not exposed to language during this time, he or she may never acquire it or may suffer severe language impairments.

Data-driven process A process that is driven entirely by the physical characteristics of a stimulus and fails to take into account the larger context or meaning.

Decay The loss of information over time from memory.

Declarative knowledge Knowledge that represents facts.

Declarative memory Memory for knowledge about facts and events. It is demonstrated by saying and occurs with conscious recall.

Declaratives A type of speech where the utterance itself is the action.

Deductive reasoning The application of the rules of logic to statements about the world. If the premise statements are correct, than the concluding statement must be as well.

Deep representations A type of image structure that consists of information in long-term memory that is used to generate a surface representation.

Deep structure The underlying meaning of a sentence that remains constant, regardless of the specific form in which it is expressed.

Defense mechanisms Constructs of the ego that reduce or redirect anxiety in various ways.

Degree of fan A concept that is related to many others has a high degree of fan. The node representing the concept will subsequently have many links radiating outward from it to other nodes.

Deliberative behavior Actions that are comprised of previously developed behaviors that are integrated into a resultant behavior and may include external and internal stimuli.

Deliberative/Reactive Paradigm (Hybrid) A robotic architecture that employs a combination of reactive behaviors and planning or other cognitive components.

Demand characteristics When participants in a study do what the stimuli, task, or experimental situation seems to demand of them.

Dendrites Branching protrusions of the neuron that receive messages from other neurons.

Dependent variable A factor measured or observed by the experimenter to see if a change has taken place.

Depth-first search A methodology for exploring a knowledge base in which information is examined in an increasingly specialized manner.

Determinism The view that all physical events are caused or determined by the sum total of all prior events.

Dichotic listening task A task used to study selective attention. It requires a participant wearing headphones to listen to two different messages played over each ear simultaneously while paying attention to only one.

Digital representation A form of representation in which information is coded in a discrete way with set values.

Directives A type of speech in which a command is issued.

Displacement The ability of language to refer to something that is removed in space or time.

Distributed artificial intelligence A field of artificial intelligence in which individual computing agents are linked together and communicate with one another cooperatively. If the agents have specialized computing strategies, then they can solve problems in tandem that neither one of them could alone.

Distributed coding/representation A coding in which an object or feature is represented by a pattern of activation among a group of cells or nodes.

Distributed cognition The view that cognition can be considered as information processing both within and between individuals.

Divided attention A form of attention that can be split or divided among several alternative information sources.

Domain-general mechanisms Those that can operate on any type of information.

Domain-specific mechanisms Those that are tuned to perform special operations on only a certain type of information.

Dorsal A direction indicating toward the top.

Dorsal visual pathway A pathway that travels upward to the parietal lobe, where information about motion and location is extracted.

Dual-code hypothesis The idea that the mind collectively uses both digital/symbolic and image representations.

Dualism The belief that the mental and physical consist of different states or substances.

Duration How long a sensation persists.

Dynamic The characteristic that languages are constantly changing as new words are added and grammatical rules altered.

Early selection model Models of attention where information is selected early, based on physical stimulus characteristics.

Echoic memory Auditory sensory memory.

Ecological models of cognition Models of cognitive processes that are acquired through learning.

Ecological perception The theory that perception is immediate and direct, not requiring representation or computation.

Ego The aspect of mind that balances the competing demands of the id and superego. It operates on the reality principle.

Electrical stimulation A method where an electrical current is passed through a bipolar electrode causing the neurons in a localized area of brain tissue to become active. The resulting activity and behavior are observed.

Electroencephalogram (EEG) A measure of the brain's gross electrical action.

Embodiment The idea that cognitive processes exist within and interact with an external environment through sensory inputs and motor outputs.

Emergent property A property that is realized through the interaction of a system's parts.

Empiricism The view that knowledge is acquired through experience.

Emulator A software program that allows one computer to mimic a second computer for the purposes of running the programs that are compatible with the second computer.

Encoding The name of the process by which information is taken in and converted into a usable mental form.

Enduring disposition Refers to automatic influences on where attention gets directed.

Engram A physical change in a specific location of the brain associated with learning.

Entity model of causation A theory of causality that says entities with specific identities are the cause of actions.

Environment of evolutionary adaptation (EEA) The period of time during which many human psychological mechanisms are believed to have evolved.

Environmental dependency syndrome A symptom of executive dysfunction in which an environmental stimulus can automatically trigger an associated behavior.

Epilepsy A disorder in which neurons fire uncontrollably, producing muscle spasms and seizures.

Epiphenomenalism The school that allows the physical to cause the mental but prohibits causation in the other direction.

Episodic memory A type of declarative memory containing knowledge of episodes or personally experienced events.

Epistemology The branch of philosophy devoted to the study of knowledge.

Equipotentiality The principle that many different parts of the brain seem to participate in memory storage.

Error signal In artificial neural networks, the difference between the actual and observed output. The error signal is used to alter connection strengths and train the network.

Ethology Study of animal behavior.

Event-related potentials (ERPs) An EEG recording of brain activity in response to a particular event.

Evolutionary computing (EC) A collection of computational methods modeled on the principles of biological evolution.

Evolutionary psychology (EP) A field of psychology that studies how evolutionary forces have shaped human mental capacities.

Evolved psychological mechanism A cognitive process that has evolved to solve a specific adaptive problem.

Exaptation or **neutral drift** Random mutations that produce new genes with little or no consequence for reproduction. They can spread in a population and at some point may assume a new adaptive function to a subsequent environmental change.

Executive control system A system in the working memory model whose function is to initiate and control ongoing processes.

Executive dysfunction A disorder characterized by a broad number of deficits including a difficulty in performing goal-directed behaviors as a result of frontal lobe damage.

Executive function Cognitive operations involved in planning, sequencing of behavior, flexible use of information, and goal attainment.

Exhaustive search When a memory search process continues matching all the way to the end of a list, even if the target has already been located.

Experimental group In an experiment, the group receiving the independent variable.

Expert system A software program that is dedicated to solving problems and providing "expert quality" advice to users.

Explanatory gap The gulf between an objective and subjective description of mental phenomena.

Explicit response A response such as a motor action that is directly observable.

Expressives A type of speech that describes the psychological states of the speaker.

Extensity The extent to which a sensation fills or occupies space.

Fallacies A fundamental misunderstanding of a statistical rule that can result from applying a heuristic.

Feed-forward network In an artificial neural network, flow of activation that is in one direction only, forward from units in an input layer to units in other layers.

Field theory The idea in physics that objects are acted on by forces in fields.

Fissure A large spacing or separation between two areas of brain tissue.

Focused attention stage A later stage in pattern recognition that requires concentrated attention under voluntary control.

Formal symbol manipulator A system that operates on contentless symbols based on a set of rules.

Formula (assertion) A database or internal computer representation of a fact.

Frame A collection of information that represents a simple concept. It includes a place to store attributes of the concept (slot) and a place for actions that can be applied to the concept (procedure).

Free association What happens when one is presented with an idea and asked to produce whatever related ideas come to mind without censoring or inhibiting.

Frontal lobe Located on the anterior portion of the cerebral hemispheres. It is implicated in problem-solving and motor behavior.

Functional kinds Things that are distinguished by their actions or tendencies.

Functional magnetic resonance imaging A variation of MRI that is used to show changes in brain activity over time.

Functionalism The view that mental states are not just physical states but the functioning or operation of those physical states.

Fundamental utilities of consciousness The role consciousness plays in the survival of the organism.

Fusiform face area (FFA) Pictures of faces activate cells in this area, located in the temporal lobe.

Fuzzy logic A methodology for translating uncertain knowledge into equivalent, but precise, linguistic representation and drawing conclusions on which to act.

Gambler's fallacy The belief that probability outcomes are not independent, that the probability of an event can be influenced by its past history.

General-purpose processor The view that the mind can solve any type of problem equally well. It is based on the notion that a problem is solved by context-independent symbol representation and computation.

Generalization The ability to apply a learned rule to a novel situation.

Generalized delta rule or **back-propagation** A way of training a network by repeatedly applying error feedback to alter connection strengths. The feedback is the difference between the actual and desired output.

Generative The property of language whereby symbolic elements can be combined to create a large number of meanings.

Geon A basic volumetric shape, such as a cube or a cylinder, that may be used in recognizing an object.

Gestalt The idea that a whole is more than just the sum of its parts.

Goal The desired end state or solution to a problem.

Graceful degradation A slow and gradual decrease in the performance of a network with increased damage.

Grammar The collection of all the rules governing a language.

Guided search A forceful or willed search for a memory item based on intelligence and reasoning.

Gyrus A ridge or fold of neural tissue.

"Hasa" link A link in a propositional network that represents property relationships.

Hebb rule States that if two connected neurons are both active simultaneously, the synapse between them will be strengthened.

Heuristic This is a mental "rule of thumb" or strategy that acts as a fast and easy way of problem solving. Heuristics are right most, but not all, of the time.

Hidden layer The second layer of a three-layer network. This is where the input layer sends its signals. It performs intermediary processing.

Hierarchical organization A semantic network where concept nodes are arranged in different levels along a continuum from abstract to concrete.

Hierarchical paradigm A highly sequential robotic architecture in which planning is a key element of its design. It is also referred to as a "top-down" approach to robotic design.

Homeostatic control Oversees the current state of the robot and makes changes in the operation as needed.

Homunculus A hypothetical "little man" inside the mind who interprets and understands information.

Hopfield-Tank networks Supervised single-layer networks where each node is usually connected to every other node. Good at regenerating clean versions of patterns and at solving optimization problems.

Horizontal plane A horizontal cut through the brain separating it into a top and bottom piece.

Hypothesis A statement derived from a theory that concerns potential outcomes in an experiment.

Iconic memory Visual sensory memory.

Id The aspect of mind that contains unconscious impulses and desires such as sex and hunger. It operates on the pleasure principle.

Idealism The belief that only mental entities are real, that the universe is essentially nonphysical.

Idealistic principle Motivates the individual to do what it considers proper.

Illumination A flash of insight, a sort of "aha" experience where a solution comes suddenly to awareness.

Image generation Occurs when the encodings in long-term memory are used to form an image in the visual buffer.

Image inspection Occurs when an individual is asked some question about an image. It consists of a number of distinct operations such as "zoom," "pan," or "scan." These operations preserve the image while extracting information from it.

Image transformation Refers to an operation performed on an image such as a mental rotation. These operations alter the image.

Immediate experience An individual's direct awareness of something.

Implementation level A "hardware" level of analysis that specifies the physical processes that carry out a computation.

Implicit response A response that occurs inside an organism and is therefore difficult to observe.

Incompatibilism The belief that free will and determinism are irreconcilable.

Incubation The period of time during which a problem is put aside. During this period there are no conscious attempts at problem solving but the unconscious mind may be attempting or have discovered a solution.

Independent variable In an experiment, a factor manipulated or altered by the experimenter to see if it will cause a change.

Inductive reasoning A form of thinking in which commonalities about specific instances in the world are noticed and used to draw conclusions.

Information gain A measure of the uncertainty of an event related to the probability of the event's occurrence.

Inheritance The process by which parents pass on some of their genetic characteristics to their offspring.

Innate Reasoning Mechanism (IRM) Similar to a reflex but operates in an open world as contrasted to a reflex that operates in a predetermined environment.

Input layer The first layer of a three-layer network that receives stimulus input and where the stimulus is represented.

Input/Output (I/O) The functional parts of a computer that enable such machines to interact with the external world.

Insight learning The apparent spontaneous understanding of relationships that produces a solution to a problem.

Intensity The strength of a sensation.

Intentionality The relationship between a representation and what it stands for.

Interactionism The belief that allows the physical and mental to causally influence one another.

Interference The inability of a network to distinguish similar patterns from one another.

Intervening variables Variables that mediate between initiating causes and a behavior.

Intraparietal sulcus Located in the parietal lobe, it may be responsible for the allocation of attentional resources and the binding of features in visual search.

Introspection The process of "looking inward" to experience and describe mental objects or states.

Ipsilateral Refers to the same side of the brain or neural structure.

"Isa" link A link in a propositional network that represents relationships of belonging.

Isomorphism The correspondence between the psychological or conscious experience on the one hand and the underlying brain experience on the other.

Knowledge-based approach A type of problem solving method in which one conceptualizes the problem and its solution in terms of symbols and transformations on the symbols. Representative of the traditional or classical approach in cognitive science.

Kohonen network An unsupervised two-layer network able to create a topological map or spatial representation of the features present in the stimulus input.

Late selection model Models of attention where information is selected late, based on semantic stimulus characteristics.

Latent learning An animal's ability to acquire behaviors through experience without being subject to any reinforcement.

Lateral A direction indicating toward the side.

Lattices Regularly arranged matrices of parts used to demonstrate the principles of perceptual organization.

Leaf (Leaves) Terminating nodes for a knowledge-base-searching algorithm.

Learning When some event causes a change in the nervous system that in turn causes a change in behavior.

Lesion study A method in neuroscience where researchers deliberately destroy brain areas in animals and examine the resulting behavioral deficits.

Lexicon A mental dictionary representing specific word features such as their sound, appearance, and meaning.

Lexigrams Geometric patterns that can be used to create a language system.

Libertarians Those who believe that people have free will and that it is not compatible with determinism.

Linguistic relativity hypothesis Also known as the Sapir-Whorf hypothesis. The strong version states that thought and language are so similar that it may be impossible to express the thoughts of one language in another. The weak version denies this but admits that the language a person speaks influences the way he or she thinks.

Linguistics The study of language.

Links The connection between each node in an artificial network.

Literal encodings Encodings that contain lists of coordinates detailing where to place points in the surface matrix to depict the represented object.

Lobes Cortical regions with different functional specializations.

Local minimum When a network is not able to perform its task because the error level drops too quickly. It is characterized by an early low point on the loss function.

Local representation Representation in the form of activation or activity in a single node in a network.

Logogen A word representation used in the logogen model.

Long-term potentiation (LTP) The enduring facilitation of synaptic transmission that occurs following activation of a synapse by repeated intense high-frequency stimulation.

Loss function The change in the error signal over a set of learning trials.

Magnetic resonance imaging (MRI) A brain imaging technique in which a patient is placed inside a tube containing a powerful magnet. Protons align themselves in this magnetic field. A radio-wave pulse is then applied to the brain or other part of the

body undergoing the scan. The radio signals are bounced back and picked up by a detector unit.

Maximal onset principle The principle that consonants usually precede vowels and typically form the onset of syllabic groupings. A feature found in all languages.

Means-ends analysis A reasoning paradigm in which the intelligence notes the current state of a system and chooses some action that will reduce the difference between the current state and a specified goal state.

Medial A direction indicating toward the middle.

Mediate experiences Experiences that come from mental reflection about an object.

Memory The capacity to retain information over time; also the part of a computer that stores the instructions to be executed, partial computational results, and final results.

Mental operations How a mental process operates, what it accomplishes, and under what conditions it occurs.

Metacognition Any process that monitors, regulates, or controls a cognitive process.

Metaphysics The branch of philosophy that examines the nature of reality.

Mimetic skill The ability to represent knowledge through voluntary motor acts.

Mission planner A robotic component that receives a mission assignment from a human—or initiates such a mission of its own—and translates the requirements of the task into terms that other parts of the robotic software understand.

Modal memory model An early model of memory showing how information is transferred between sensory, short-term, and long-term memory.

Modularity of mind A theory that states the mind is made up of innate and functionally distinct modules.

Molecular drive Occurs when a copy of a gene mutates and serves an adaptive function even though it was not selected for.

Momentary intention Refers to the conscious decision to pay attention to something.

Monism The belief that the mental and physical consist of a single type of state or substance.

Morphemes Units of spoken language that have meaning.

Morphology The rules governing the word structure of a language.

Mosaic model of development A model in which brain mechanisms are almost entirely determined by genes, operate quickly, and the parts of the system develop independently.

Multiagent system A collection of specialized interacting agents.

Multilayer network A network that has two or more layers of nodes.

Multimode model of attention The view that selection can be based on multiple stimulus characteristics such as the physical and semantic.

Multiple-unit recording A brain recording technique where an electrode is used to measure the collective electrical activity of a group of neurons.

Narrative thought This is a type of thought that is subjective and holistic and is employed more in the arts and humanities.

Nativism A belief that a significant body of knowledge is innate or "built into" an organism.

Natural languages Languages that have evolved in humans and are used by them. Examples include English, Spanish, and French.

Nature-nurture debate A controversy that centers on the relative contribution of biology and experience in determining any particular capacity.

Navigator This software process creates a path (e.g., set of points defining a straight line) that a robot is intended to follow.

Neural Darwinism A theory that applies evolutionary processes to neural learning. In this view, neuronal circuits that are active survive and those that are not die out.

Neural synchrony A theory that suggests an object is represented by the joined and coordinated activity of a constellation of distributed cells. It is one solution to the binding problem.

Neurons Cells that receive and transmit electrical impulses. They form the basic units of the nervous system.

Neuroscience The study of nervous system anatomy and physiology.

Neurotransmitters Molecules released from the terminal button that travel across the synaptic cleft and attach to receptor sites on the surface of another cell.

Node A representation of a concept or element of the world.

Nonreductive physicalists Physicalists who believe physical processes can give rise to emergent and irreducible mental phenomena.

Object-centered description An object representation that is described relative to the object itself and can be recognized from a variety of perspectives.

Object constancy The ability to recognize an object even though it is hardly ever viewed from the same perspective twice.

Object link A link in a propositional network that denotes the object or thing to which an action is directed.

Occipital lobe Located on the posterior portion of the cerebral hemispheres, it mediates visual processing.

One-word stage A stage in language development where words are first used in a symbolic and semantic fashion.

Open world paradigm References environments in which actions of intelligent agents alter objects within their domain and thus require a resulting modification of the environment.

Operant conditioning A form of learning in which reward and punishment shape the behavior of an entity (e.g., human).

Operator A process that is applied to the problem-solving situation and that transforms it.

Ordinate A level of concept category organization of moderate specificity.

Output layer The third layer of a three-layer network. This generates a representation of the response based on inputs from the hidden layer.

Pandemonium model A feature detection model of pattern recognition in which "demons" are used to represent different aspects of a stimulus.

Paradigm A philosophy or approach for developing theories for, analyzing, and evaluating a class of problems.

Paradigmatic thought This is a type of thought that is objective and analytical and is implemented more often in the sciences.

Parallel distributed processing (PDP) A type of information processing where large numbers of computing units perform their calculations simultaneously. Computational units can receive and process multiple inputs and transmit multiple outputs.

Parallel memory search A memory search in which all the items in a list are matched against a target all at once.

Parallel search Visual search where all the items in the display can be matched for the target at once.

Parallelism The belief that the mind and body are distinct and isolated from each other.

Parietal lobe Located dorsally, it is implicated in somatosensory, attention, and spatial processing.

Partial-report condition A condition in iconic memory studies in which participants are cued to remember only one row of letters in the display.

Pattern recognition The ability to identify objects in the environment.

Perception The process by which we gather and interpret information from the outside world via the senses; within a mechanical environment, perception refers to the process of scanning an environment and converting the resultant information into a set of abstractions comprising objects, features, and relationships.

Perceptron Artificial neural networks that detect and recognize information about the world, store this information, and use it in some fashion. They are characterized by the ability to learn from experience and can alter their connection strengths.

Perceptual categorization deficit A difficulty in recognizing objects when they are viewed from an unusual angle or are lit unevenly.

Pessimists Theorists who argue that moral free will is impossible to prove.

Phase sequence According to Hebb, a group of connected cell assemblies that fire together or close in time.

Phenomenal concept of mind The idea of mind as a conscious experience.

Phenomenology Refers to subjective experience rather than objective description.

Philosophy A discipline concerned with the search for wisdom and knowledge.

Phoneme The smallest unit of sound in a language.

Phonology The rules governing the sound system of a language.

Phrase structure The hierarchical organization of a sentence. Represented using a tree diagram that shows the relationships between words.

Phrase structure grammar The grammar that governs the use of phrase structures.

Physical kinds Things that are identified by their material composition only.

Physicalism Also known as materialism; the belief that the universe is entirely physical.

Pilot This software process determines the actions that a robot must complete to follow the path specified by the Navigator.

Plan A robotic primitive element encompassing the corresponding human attributes of reasoning and cognition.

Pleasure principle Motivates the immediate satisfaction of desires.

Pop-out A perceptual phenomenon that occurs when a target is easy to find because it differs from the surrounding distracters along a single stimulus dimension such as color.

Positron emission tomography (PET) A brain imaging technique in which blood flow through the brain is measured while a participant carries out a cognitive task. This is accomplished using radioactive isotopes attached to tracer molecules such as glucose or oxygen. Brain areas that are more active will use more of the tracer molecules. This increase in resulting activity can then be detected.

Posterior A direction indicating toward the back.

Pragmatic analysis The last stage of natural language processing in which a complete meaning for a sentence is determined by applying contextual information.

Pragmatics The social rules underlying language use and the strategies used by speakers to make themselves clear.

Pragnanz Meaning "good figure," the Gestalt principle stating parts that are simple will group together.

Preattentive stage An early stage in pattern recognition that happens automatically and effortlessly. It does not require conscious focused attention.

Preconscious That aspect of mind that we can bring into awareness with effort.

Predicate An assertion of a fact about one or more entities or subjects.

Predicate calculus A general system of logic that accurately expresses a large variety of assertions and modes of reasoning.

Prediction The scientific view that if all states of a physical system are known, the actions that occur in them can be anticipated.

Preparation An understanding of and preliminary attempts to solve a problem.

Primacy effect The phenomenon that more words are remembered at the beginning of a list.

Primary motor cortex Located on the precentral gyrus, it contains a spatial representation or map of different body areas used to control motor behavior.

Primary somatosensory cortex Located on the postcentral gyrus, it contains a spatial representation or map of different body areas used to process sensory information from the body surface.

Priming Occurs when processing of a stimulus is facilitated by the prior presentation of a related stimulus.

Principles of perceptual organization Ways in which visual parts group together to form objects.

Proactive interference Occurs when information that is learned earlier interferes with remembering information learned later.

Problem-solving A type of behavior that involves deciding that a problem exists and considering one or several solutions to the problem. Special-purpose solutions depend on the circumstances of a particular task. General-purpose solutions can be applied to a broad variety of problems.

Problem space The initial, intermediate, and goal states of the problem.

Procedural knowledge Knowledge that represents skills.

Procedural memory Memory for skill knowledge. It is demonstrated by doing and occurs without conscious recall.

Process model A diagrammatic model that represents the way human information is processed. In the model, boxes are used to designate each stage or step in an information-processing sequence.

Production Rule (Conditional statements) Propositions or predicates whose truth or falsity can be determined.

Program A series of instructions that an intelligence devises for the computer to execute.

Proposition A statement or assertion typically posed in the form of a simple sentence.

Propositional encodings Abstract, language-like representations, similar to declarative statements. They contain information about an object's parts, its location, and size.

Propositional hypothesis Mental representations that take the form of abstract, sentence-like structures.

Prosopagnosia An inability to recognize faces, despite the capacity to recognize other types of visual stimuli and the presence of generally intact intellectual functioning.

Prototypes A generic or idealized representation of a conceptual category.

Proximity The principle stating parts that are close to one another in the visual field are perceived as a whole.

Psychoanalytic psychology The view that the mind is made up of distinct components, each competing with one another and vying for control of behavior.

Psychological concept of mind The idea of mind as mental states that cause and explain behavior.

Psychological inertia A symptom of executive dysfunction characterized by listlessness and an inability to stop some action once it is started.

Psychologist's fallacy The idea that one person's subjective response to a perception does not guarantee that the same response will exist in the mind of anyone else who experiences the same perception.

Psychology The scientific study of mind and behavior.

Psychophysical relations The relation between the psychological mind and the physical body.

Punishment Any consequence that decreases the frequency of a preceding behavior.

Qualia The felt or experienced character of mental states.

Quality The characteristic that distinguishes sensations from each other.

Rationalism The belief in the existence of innate ideas.

Raw primal sketch An image represented in terms of its distribution of intensity values or areas of light and dark.

Reactive paradigm (Subsumption) A robotic architecture characterized by direct connection between perceptive elements (sensory information) and actuating elements (sense-action processes).

Reactive responses Learned, consolidated behaviors that are executed without conscious thoughts.

Reagent A substance added to a mixture to produce a particular chemical reaction.

Reality principle Motivates one to act in a rational and pragmatic fashion.

Reasoning The ability to draw inferences appropriate to the facts and circumstances of the situation.

Recency effect The phenomenon that more words are remembered at the end of a list.

Receptors Structures on the surface of a neuron to which neurotransmitters attach.

Reciprocal altruism A characteristic of early human society involving the sharing of hard-won resources between group members.

Reconstructive memory When recall of an item from memory is based on guided search and subject to bias by subsequent information.

Recurrent networks This is when information in a network can flow in two directions, both forward and backward.

Reductionism The idea that an understanding of parts can completely account for the behavior of a whole.

Reductive physicalism A school of physicalism that states that mental phenomena can be explained by the brain's physical makeup and processes.

Referent The thing or things in the external world that a representation stands for.

Reflex A behavior in which a stimulus triggers sensory neurons that activate intermediary neurons. These in turn activate motor neurons, causing a simple motor response.

Reflexive responses Actions that last as long as the stimulus that produced them and whose magnitude is proportional to the intensity of the stimulus. Fixed action responses continue for a longer time duration than the stimulus.

Regulatory model of development A model in which brain mechanisms are only partially determined by genes, operate more slowly, and the parts of the system develop interdependently.

Rehearsal The repeated activation of a circuit that increases the strength of the connections between its nodes.

Reinforcement Any event that increases the frequency of a preceding response.

Relation link A link in a propositional network specifying the type of relation between agent and object.

Replication The scientific view that if all states of a physical system are known, the actions that occur in them can be reproduced.

Representativeness heuristic The tendency to judge an item based on its perceived similarity to a category label.

Repression A defense mechanism that banishes anxiety-arousing thoughts and feelings from consciousness.

Reticular activating system (RAS) A network of about 100 nuclei that control the brain's overall arousal and alertness level.

Retrieval The act of accessing needed data from memory and making it available for use.

Retrieval cues A phenomenon in which an item related to one that was memorized can lead to successful recall.

Retroactive interference Occurs when information that is learned later interferes with remembering information learned earlier.

Retrograde amnesia An inability to remember information acquired prior to some traumatic event.

Reverse engineering The process of starting with an end-product and analyzing it to determine its intended function.

Robot A mechanical agent that can function autonomously.

Roboticist A scientist whose research interest is the study of robots.

Root node The starting conceptual point for searching a knowledge base.

Rote learning Learning characterized by a direct association between stimulus and response.

Sagittal plane A vertical cut through the brain separating it into a left and right half.

Schema Knowledge of how to act as well as the computational processes by which to accomplish the activity (algorithm).

Scientific method A process of studying natural phenomena that involves observation and the testing of hypotheses through the use of experiments.

Script A type of schema that acts as a stored framework or body of knowledge about some topic.

Selection A change in environmental conditions that results in differential inheritance of traits in a population.

Selective attention A form of attention that can be focused onto one source of information and away from another.

Self-terminating search A memory search process that stops as soon as a positive match between a list item and target occurs.

Semantic The meaning of a representation.

Semantic analysis The third step of natural language understanding where the sentence structure and the meaning of the words are used to derive a partial representation of the meaning of a sentence.

Semantic code A memory code based on the inherent meaning of a stimulus.

Semantic memory A type of declarative memory containing knowledge of facts.

Semantic network A network where each node has a specific meaning and therefore employs local representation of concepts.

Semantics The rules for deriving meaning in a language.

Sense A robotic primitive includes that part of a robotic system that converts elements of an environment into information that is made available to other parts of the robotic system.

Sensory memory A short-term repository for incoming sensory information.

Sentence verification A procedure in which participants judge the truth or falsity of sentences by pushing one of two buttons.

Serial memory search A memory search in which each item in a list is matched against a target one at time.

Serial processors A type of information processing where one computation is performed at a time. The results of one computation typically serve as the input for the next.

Serial search Visual search where each item in a display must be scrutinized one after the other until the target is located.

Sexual division of labor A division of labor between the sexes that is believed to have existed in early human societies where males hunted and females gathered.

Sexual selection A differential inheritance of traits based on mate selection and competition instead of a change in the environment.

Shared representations Representations that can be processed by more than one user. Language is a common medium of shared representations.

Shiftability The idea that attention can be shifted back and forth between different sources.

Similarity The perceptual principle stating parts that are similar in lightness, color, shape, or some other characteristic group together.

Simple ideas Ideas derived through sensory input or simple processes of reflection.

Single-cell recording A brain recording technique where a fine microelectrode is inserted either into a single neuron or into the extracellular fluid adjacent to it. Changes in that cell's electrical conductivity or its rate of firing are then measured.

Single-layer A network with only one layer of nodes.

Solipsism The belief that the universe exists only in one's mind.

Spandrel An architectural feature formed by the triangular space between the exterior curve of an arch and the rectangular frame enclosing it. In evolution, it is used to designate a byproduct of an adaptation that may serve a useful purpose.

Specificity coding A coding in which activity in a single neuron represents an object or feature.

Speech recognition The first step in the process of natural language processing by which the acoustic speech signal is analyzed to yield the sequence of spoken words.

Speech spectrogram A plot that displays the different frequency components of speech and how they change over time.

Spreading activation The activity that spreads outward from nodes along links to activate other nodes.

Stability-plasticity dilemma A scenario where a network should be plastic to store novel input patterns. At the same time it should be stable to protect previously encoded patterns from being erased.

Stimulus error Confusing our true experience of an object for a description of the object based on language and past experience.

Storage Information that is represented but not currently activated for use.

Stream of consciousness The notion that the mind is a process undergoing continuous flow or change.

Structuralism A theory that emphasizes studying the structure or basic elements of mind and how they combine.

Structured A property of language in which rules specify how symbols can be combined.

Subgoal An intermediate goal along the route to eventual solution of a problem.

Sublimation A defense mechanism that transforms unacceptable impulses into socially valued motivations.

Subordinate The most concrete or specific form of conceptual category organization.

Substantive thought Occurs when the mind slows down, perhaps through the focusing of attention.

Sulcus A smaller spacing or separation between two areas of brain tissue.

Superego The aspect of mind that is responsible for our ethical sense of right and wrong. It operates on the idealistic principle.

Superior colliculus A brain area responsible for the moving of visual attention from one object or position in space to another.

Superordinate The most abstract form of conceptual category organization that encompasses all examples of the concept.

Supervised networks These are networks that are presented with target answers for each pattern they receive as input. The network "knows" what the right answer is on each training trial.

Supervisory Attentional System (SAS) A set of schemas that govern only nonroutine actions that require controlled attentional processes. These schemas are designed to be more general purpose and applicable to a wide variety of different problem types.

Surface representation A quasi-pictorial representation that occurs in a spatial medium. It depicts an object or scene and underlies our experience of imagery.

Surface structure The organization of a sentence in the form that it is expressed. How a sentence would be heard if it were spoken or read if it were written.

Sustainability The idea that attention can be maintained over time.

Symbol A representation that stands for something else.

Synaptic cleft The space in between neurons across which neurotransmitter molecules travel. Typically found between the terminal button of one cell and the dendrite of another.

Synaptic plasticity A change in the structure or biochemistry of a synapse that occurs during learning.

Syntactic analysis The second step of natural language processing where the word sequence is analyzed using knowledge of the language's grammar. This produces sentence structure.

Syntax The rules that govern the arrangement of words together in sentences.

Teacher, The A corrective mechanism that compares actual to desired output and generates an error signal.

Template An internal mental representation of a stimulus to which an image generated from an external stimulus is matched.

Temporal lobe Located laterally on each hemisphere. It mediates auditory processing, pattern recognition, and language comprehension.

Terminal button A bulb-like structure found at the end of the axon. It releases neurotransmitters.

Thalamus This brain structure serves as a relay center, forwarding incoming messages from the different senses to parts of the cortex specialized for processing them.

Theory A set of statements that organizes facts and aids in understanding how the world works.

Theory of natural selection The theory proposed by Charles Darwin that accounts for changes in animal species over time. It involves species variability, inheritance of traits through reproduction, and selection due to environmental change.

3-D sketch A three-dimensional image representation in which object parts are linked together by axes of symmetry and elongation.

Threshold The minimum amount of activation required to produce conscious awareness of a stimulus.

Threshold of excitation The minimal change in a neuron's normal resting electrical state that will initiate the creation of an electrical signal (action potential).

Token Referring to a specific instance of a category.

Token (instance) An identification symbol that is useful for symbolic manipulation of knowledge.

T.O.T. phenomenon This is an acronym for "tip of the tongue" in which one feels familiarity with an item but cannot quite recall it.

Transformational grammar A set of rules for modifying a sentence into a closely related one.

Transitive thought The flow of thought that occurs during less focused and more associative forms of reasoning.

Trial-and-error learning Learning that proceeds from random responses to a behavior that satisfies the requirements of the task.

Tridimensional theory of feeling A theory stating that all feelings can be characterized by three dimensions: pleasure-displeasure, tension-relaxation, and excitement-depression.

2 1/2-D sketch An image representation that includes information about surfaces and layout.

Two-word stage A stage in language development in which children first produce two-word utterances.

Type Referring to an entire category.

Typicality effect The phenomenon that human participants are faster to judge stereotypical members as belonging to a category.

Unconditioned response (UCR) A response elicited by an unconditioned stimulus.

Unconditioned stimulus (UCS) A stimulus that elicits a response on its own.

Unconscious That aspect of mind of which we are completely unaware.

Universal grammar The features that are instantiated in the grammars of all natural languages.

Unsupervised networks Networks that must determine the answer on their own without any benefit of a corrective response.

Variation Refers to the differences in traits between animals.

Ventral A direction indicating toward the bottom.

Ventral visual pathway A pathway that travels downward to the temporal lobe and carries data about color and form.

Verification A stage in problem solving where an insight is confirmed and one checks to see that it produces a correct solution.

Viewer-centered description An object representation that is particular to the viewer's point of view.

Visual agnosia A deficit resulting from brain damage in which an individual has difficulty in recognizing objects visually.

Visual buffer The spatial medium of the surface representation. The buffer is a surface matrix consisting of an array of points.

Visual code A memory code based on the visual appearance of a stimulus.

Visual image A mental representation of an object or scene that preserves metric spatial information.

Visual search The task of identifying a target item located in a field filled with nontarget items or distracters.

Visuo-spatial sketchpad A system in the working memory model that is specialized for the processing of visual information.

Voluntarism A movement that viewed the mind as consisting of elements and stressed that these elements are assembled into higher-level cognitive components through the power of the will.

Wason Selection Task A task designed to measure a person's logical thinking ability. It involves applying the abstract rules of logic to a specific example.

Weights Values on links that determine the strength of the connection between nodes. They can range in value from –1.0 to +1.0.

Wernicke's aphasia A language deficit in which speech production is intact but there are impairments in comprehension.

Wernicke's area A region in the left temporal lobe that underlies language comprehension. Damage to this area results in Wernicke's aphasia.

Whole-report condition A condition in iconic memory studies in which the task is to remember the entire letter display.

Working memory A short-term memory store used to represent and transform information.

Zeigarnik effect The phenomenon of remembering more information following an interruption.

References

Ackley, D., & Littman, M. (1992). Interactions between learning and evolution. In C. Langton (Ed.), *Artificial life II* (pp. 487–509). Redwood City, CA: Addison-Wesley.

Alvarez, P., & Squire, L. R. (1994). Memory consolidation and the medial temporal lobe: A simple network model. *Proceedings of the National Academy of Science, USA, 91(15),* 7041–7045.

Anderson, J. R. (1980). *Cognitive psychology and its implications* (2nd ed.). San Francisco: W. H. Freeman.

Anderson, J. R. (1980). Concepts, propositions, and schemata: What are the cognitive units? *Symposium on Motivation, 28,* 121–162.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2,* 396–408.

Angell, J. R. (1907). The province of functional psychology. *Psychological Review, 14,* 61–91.

Arbib, M. (1981). Perceptual structures and distributed motor control. In V. B. Brooks (Ed.), *Handbook of physiology: The nervous system* (Vol. 2, pp. 1449–1465). Baltimore: American Physiological Society.

Arkin, R. (1998). *Behavior-based robotics.* Cambridge, MA: MIT Press.

Arkin, R. C., Riseman, E. M., & Hansen, A. (1987). AuRA: An architecture for vision-based robot navigation. *Proceedings of the DARPA Image Understanding Workshop,* 413–417.

Ashcraft, M. H. (2002). *Cognition* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225,* 82–90.

Au, T. K. (1983). Chinese and English counterfactuals: The Sapir-Whorf hypothesis revisited. *Cognition, 15(1–3),* 155–187.

Au, T. K. (1984). Counterfactuals: In reply to Alfred Bloom. *Cognition, 17(3),* 289–302.

Averbach, E., & Sperling, G. (1961). Short term storage and information in vision. In C. Cherry (Ed.), *Information theory* (pp. 196–211). London: Butterworth.

Baddeley, A. D. (1986). *Working memory.* Oxford, UK: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science, 255,* 556–559.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113,* 1–29.

Banich, M. T. (2004). *Cognitive neuroscience and neuropsychology.* Boston: Houghton Mifflin.

Barr, A., & Feigenbaum, E. A. (Eds.). (1982). *The handbook of artificial intelligence* (Vol. 2). Stanford, CA: Heuris Tech Press.

Barto, A. G., Sutton, R. S., & Anderson, C. W. (1983). Neuronlike adaptive elements that can solve difficult learning control problems. *IEEE Transactions on Systems, Man, & Cybernetics, 13(5),* 834–846.

Bavelier, D., Corina, D., Jezzard, P., Clark, V., Karni, A., Lalwani, A., et al. (1998). Hemispheric specialization for English and ASL: Left invariance–right variability. *Neuroreport, 9(7),* 1537–1542.

Baylis, G. C., Rolls, E. T., & Leonard, C. M. (1985). Selectivity between faces in the responses of a population of neurons in the cortex in the superior temporal sulcus of the monkey. *Brain Research, 342,* 91–102.

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks.* Cambridge, MA: Basil Blackwell.

Beer, R. D. (1990). *Intelligence as adaptive behavior: An experiment in computational neuroethology.* New York: Academic Press.

Bellugi, U., Wang, P. P., & Jernigan, T. L. (1994). Williams syndrome: An unusual neuropsychological profile. In S. H. Broman & J. Grafman (Eds.), *Atypical cognitive deficits in developmental disorders: Implications for brain function* (pp. 23–56). Hillsdale, NJ: Erlbaum.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution.* Berkeley: University of California Press.

Besner, D., & Stolz, J. A. (1999). The Stroop effect and single letter coloring: What replicates and what doesn't. *Psychonomic Bulletin & Review, 8(4),* 858.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94,* 115–147.

Bishop, D. V. M. (1997). *Uncommon understanding: Development and disorders of language comprehension in children.* Hove, UK: Psychology Press.

Black, M. (1971) The elusiveness of sets. *The Review of Metaphysics, 24,* 614–636.

Blackmore, S. (2004). *Consciousness: An introduction.* Oxford, UK: Oxford University Press.

Block, N. (1981). Introduction: What is the issue? In N. Block (Ed.), *Imagery* (pp. 1–18). Cambridge, MA: MIT Press.

Block, N. (1981). Psychologism and behaviorism. *Philosophical Review, 90,* 5–43.

Block, N. (1995). The mind as the software of the brain. In D. Osherson, L. Gleitman, S. Kosslyn, E. Smith & S. Sternberg (Eds.), *An Invitation to cognitive science.* Cambridge, MA: MIT Press.

Bloom, A. H. (1981). *The linguistic shaping of thought: A study in the impact of language on thinking in China and the West.* Mahwah, NJ: Erlbaum.

Bloom, P. (Ed.). (1994). *Language acquisition.* Cambridge, MA: MIT Press.

Boden, M. (1990). Escaping from the Chinese room. In M. Boden (Ed.), *The philosophy of artificial intelligence* (pp. 88–104). Oxford, UK: Oxford University Press.

Bond, A. H., & Gasser, L. (Eds.). (1988). *Readings in distributed artificial intelligence.* San Francisco: Kaufmann.

Bond, M. H. (1988). Finding universal dimensions of individual variation in multi-cultural studies of values: The Rokeach and Chinese values surveys. *Journal of Personality and Social Psychology, 55,* 1009–1015.

Boster, J., & D'Andrade, R. (1989). Natural and human sources of cross-cultural agreement in ornithological classification. *American Anthropologist, 91,* 132–142.

Bowlby, J. (1967). *Attachment and loss: Vol. 1. Attachment.* New York: Basic Books.

Braitenberg, V. (1984). *Vehicles: Experiments in synthetic psychology.* Cambridge, MA: MIT Press.

Braspenning, P. J. (1995). Neural cognodynamics. In P. J. Braspenning, F. Thuijsman & A. J. M. M. Weijters (Eds.), *Artificial neural networks: An introduction to ANN theory and practice* (pp. 247–272). Berlin: Springer-Verlag.

Bringsjord, S. (1994). Could, how could we tell if, and should—Androids have inner lives. In K. Ford, C. Glymour & P. Hayes (Eds.), *Android epistemology* (pp. 93–122). Cambridge, MA: MIT Press.

Broadbent, D. E. (1958). *Perception and communication.* London: Pergamon.

Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation, 1,* 1–10.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence, 47,* 139–159.

Brooks, R. A. (1999). *Cambrian intelligence.* Cambridge, MA: MIT Press.

Brooks, R. A. (2002). *Flesh and machines: How robots will change us.* New York: Random House.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (4th ed., pp. 77–166). New York: John Wiley.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin, 109(2),* 204–223.

Brown, J. A. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10*, 12-21.

Bruce, C., Desimone, R, & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in the superior temporal sulcus of the macaque. *Journal of Neurophysiology, 46*, 369-384.

Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.

Burton, A. M., Young, A. W., Bruce, V., Johnston, R. A., & Ellis, A. W. (1991). Understanding covert recognition. *Cognition, 39,* 129-166.

Buss, D. M. (1989). Conflict between the sexes: Strategic interference and the evocation of anger and upset. *Journal of Personality and Social Psychology, 56,* 735–747.

Buss, D. M. (1999). *Evolutionary psychology: The new science of the mind*. Boston: Allyn & Bacon.

Caramazza, A., & Shelton, J. (1998). Domain-specific knowledge systems in the brain: The animate/inanimate distinction. *Journal of Cognitive Neuroscience, 10,* 1–34.

Carlson, N. R. (2001). *Physiology of behavior*. Boston: Allyn and Bacon.

Carroll, J. B. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge, MA: MIT Press.

Carpenter, G. A., & Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *Computer, 21,* 77–90.

Carr, H. A. (1925). *Psychology*. New York: Longmans, Green.

Cawsey, A. (1998). *The essence of artificial intelligence*. Harlow, UK: Prentice Hall.

Chalmers, D. (1996). *The conscious mind*. Oxford, UK: Oxford University Press.

Chalmers, D. J. (1997). Availability: The cognitive basis of experience? In N. Block, O. Flanagan & G. Guzeldere (Eds.), *The nature of consciousness*. Cambridge, MA: MIT Press.

Chapanis, A. (1965). Color names for color space. *American Scientist, 53,* 327–346.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America, 25,* 975–979.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Chomsky, N. (1959). A review of Skinner's *Verbal Behavior*. *Language, 35,* 26–58.

Chomsky, N. (1968). *Language and mind*. New York: Harcourt Brace & World.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul: A philosophical journey into the brain*. Cambridge, MA: MIT Press.

Churchland, P. S., Koch, C., & Sejnowski, T. J. (1990). What is computational neuroscience? In E. L. Schwartz (Ed.), *Computational neurosciences* (pp. 38–45). Cambridge, MA: MIT Press.

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science*. New York: Oxford University Press.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.

Clark, V. P., Keil, K., Maisog, J. M., Courtney, S. M., Ungerleider, L. G., & Haxby, J. V. (1996). Functional magnetic resonance imaging of human visual cortex during face matching: A comparison with positron emission tomography. *NeuroImage, 4*, 1-15.

Colby, C. L., Duhamel, J.-R., & Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of Neurophysiology, 76,* 2841–2852.

Colby, K. M. (1975). *Artificial paranoia: A computer simulation of paranoid processes*. New York: Pergamon Press.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior, 8,* 240–247.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology, 55,* 75–84.

Cook, V. J. (1988). *Chomsky's universal grammar: An introduction*. Cambridge, MA: Basil Blackwell.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75–176). New York: Academic Press.

Corkin, S. (1984). Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental findings in H. M. *Seminars in Neurology, 4,* 249–259.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind* (pp. 163–228). New York: Oxford University Press.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience, 18,* 7426–7435.

Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature, 375,* 121–123.

Crystal, D. (1987). *The Cambridge encyclopedia of language*. Cambridge, MA: Cambridge University Press.

Cummins, D. D. (1999, June). *Early emergence of cheater detection in human development*. Paper presented at the annual meeting of the Human Behavior and Evolution Society, Salt Lake City, UT.

Damasio, A. (1994). *Descartes' error: Emotion, reason and the human brain*. New York: Putnam.

Damasio, A. R., & Tranel, D. (1993). Verbs and nouns are retrieved from separate neural systems. *Proceedings of the National Academy of Sciences, 90,* 4957–4960.

Darwin, C. (1859). *On the origin of species by means of natural selection*. London: John Murray.

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology, 3,* 255–267.

Davies, I. R. L., & Corbett, G. G. (1997). A cross-cultural study of colour grouping: Evidence for weak linguistic relativity. *British Journal of Psychology, 88(3),* 493–517.

Davies, I. R. L., Sowden, P. T., Jerrett, D. T., Jerrett, T., & Corbett, G. G. (1998). A cross-cultural study of English and Setswana speakers on a colour triads task: A test of the Sapir-Whorf hypothesis. *British Journal of Psychology, 89(1),* 1–15.

Dawson, M. R. W. (1998). *Understanding cognitive science.* Malden, MA: Blackwell.

Dayhoff, J. (1990). *Neural network architectures: An introduction.* New York: Van Nostrand Reinhold.

Dennett, D. (1991). *Consciousness explained.* Boston: Little Brown & Company.

Dennett, D. (1998). The practical requirements for making a conscious robot. In D. C. Dennett, *Brainchildren: Essays on designing minds* (pp. 154–170). Cambridge, MA: MIT Press.

Dennett, D. C. (1981). Two approaches to mental images. In N. Block (Ed.), *Imagery* (pp. 87–107). Cambridge, MA: MIT Press.

Deregowski, J. B. (1989). Real space and represented space: Cross-cultural perspectives. *Behavioral and Brain Sciences, 12,* 51–119.

Desimone, R., Albright, T. D., Gross, C. D., & Bruce, C. (1984). Stimulus-selective responses of inferior temporal neurons in the macaque. *Journal of Neuroscience, 4,* 2051–2062.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70,* 80–90.

Domany, E., Van Hemmen, J. L., & Schulten, K. (Eds.) (1994). *Models of neural networks* (Vol. 2, pp. 95–119). New York: Springer-Verlag.

Donald, Merlin. (1991). *Origins of the modern mind.* Cambridge, MA: Harvard University Press.

Donne, J. (1624). *Meditation XVII. Devotions upon emergent occasions.*

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience, 22,* 567–631.

Dover, G. (2000). Anti-Dawkins. In H. Rose & S. Rose (Eds.), *Arguments against evolutionary psychology* (pp. 55–77). New York: Harmony Books.

Dreyfus, H. L. (1992). *What computers still can't do* (3rd ed.). Cambridge, MA: MIT Press.

Dronkers, N. F., Redfern, B. B., & Ludy, C. A. (1995). Lesion localization in chronic Wernicke's aphasia. *Brain and Language 51(1),* 62–65.

Dronkers, N. F., Shapiro, J. K., Redfern, B., & Knight, R. T. (1992). The role of Broca's area in Broca's aphasia. *Journal of Clinical and Experimental Neuropsychology, 14,* 52–53.

Dunbar, R. (1996). *Grooming, gossip, and the evolution of language.* Cambridge, MA: Harvard University Press.

Duncker, K. (1945). On problem solving. *Psychological Monographs, 58,* 1–112 (Whole No. 270).

Dworkin, B. R. (1993). *Learning and physiological regulation.* University of Chicago Press.

Edelman, G. M. (1989). *Neural darwinism: The theory of neuronal group selection.* Oxford, UK: Oxford University Press.

Edwards, F. A. (1995). LTP—a structural model to explain the inconsistencies. *Trends in Neuroscience, 18*, 250-255.

Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion. In H. Wagner & A. Manstead (Eds.), *Handbook of social psychophysiology* (pp. 143–164). Chichester, UK: Wiley.

Elman, J. (1990). Finding structure in time. *Cognitive Science, 14,* 179–211.

Ember, C. R., & Ember, M. (1985). *Cultural anthropology* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Engel, A. K., König, P., Kreiter, A. K., Schillen, T. B., & Singer, W. (1992). Temporal coding in the visual cortex: New vistas on integration in the nervous system. *Trends in Neurosciences, 15,* 218–226.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95,* 307–317.

Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision.* Cambridge, MA: M.I.T. Press.

Farah, M. J. (2000). *The cognitive neuroscience of vision.* Malden, MA: Blackwell Publishers.

Fink, G. R., Dolan, R. J., Halligan, P. W., Marshall, J. C., & Firth, C. D. (1997). Space-based and object-based visual attention: shared and specific neural domains. *Brain, 120,* 2013–2028.

Finke, R. A., & Kosslyn, S. M. (1980). Mental imagery acuity in the peripheral visual field. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 126–139.

Fodor, J. A. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Forbus, K., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science, 19,* 144–205.

Fox, J. (1981). Towards a reconciliation of fuzzy logic and standard logic. *International Journal of Man-Machine Studies, 15,* 213–220.

Freedman, D. (1994). *Brainmakers.* New York: Simon & Schuster.

French, R. M. (1992). Semi-distributed representations and catastrophic forgetting in connectionist networks. *Connection Science, 4,* 365–377.

French, R. M. (2001). Catastrophic interference in connectionist networks. In *Macmillan Encyclopedia of the Cognitive Sciences.* London: Macmillan.

Frieze, I. H., Parsons, J. E, Johnson, P. B., Ruble, D. N., & Zellman, G. L. (1978). *Women and sex roles: A social psychological perspective.* New York: Norton.

Fromkin, V. (1974). The development of language in Genie: A case of language acquisition beyond the "critical period." *Brain & Language, 1(1),* 81–107.

Gabrieli, J. D. E., Desmond, J. E., Demb, J. B., & Wagner, A. D. (1996). Functional magnetic resonance imaging of semantic memory processes in the frontal lobes. *Psychological Science, 7(5),* 278–283.

Gainotti, G., & Silveri, M. C. (1996). Cognitive and anatomical locus of lesion in a patient with a category-specific semantic impairment for living beings. *Cognitive Neuropsychology, 13,* 357–389.

Gardner, H. (1974). *The shattered mind*. New York: Vintage.

Gardner, H. (1985). *The mind's new science*. New York: Basic Books.

Gardner, R. A., Gardner, B. T., & Van Cantfort, T. E. (1989). *Teaching sign language to chimpanzees*. New York: SUNY Press.

Garfield, J. L. (1995). Philosophy: Foundations of cognitive science. In N. A. Stillings, S. E. Weisler, C. H. Chase, M. H. Feinstein, J. L. Garfield & E. L. Rissland (Eds.), *Cognitive science: An introduction* (2nd ed., pp. 331–377). Cambridge, MA: MIT Press.

Gaulin, S., & FitzGerald, R. (1986). Sex differences in spatial ability: An evolutionary hypothesis and test. *American Naturalist, 127,* 74–88.

Gaulin, S., & FitzGerald, R. (1989). Sexual selection for spatial learning ability. *Animal Behavior, 37,* 322–331.

Gaulin, S. J. C., & McBurney, D. H. (2001). *Psychology: An evolutionary approach*. Upper Saddle River, NJ: Prentice Hall.

Geschwind, N. (1972). Language and the brain. *Scientific American, 226,* 76–83.

Gibson, J. J. (1986). *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306–355.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction. *Psychological Review, 102,* 684–704.

Gigerenzer, G., & Hug, K. (1992). Domain specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43,* 127–171.

Goldenberg, G., Podreka, I., Steiner, M., Willmes, K., Suess, E., & Deecke, L. (1989). Regional cerebral blood flow patterns in visual imagery. *Neuropsychologia, 27,* 641–664.

Goldman-Rakic, P. S. (1993). Working memory and the mind. *Scientific American, 267,* 110–117.

Goldstein, E. B. (2002). *Sensation and perception*. 6th Edition. Pacific Grove, CA: Wadsworth-Thompson.

Goodglass, H., & Geschwind, N. (1976). Language disorders (aphasia). In E. C. Carterete & M. P. Friedman (Eds.), *Handbook of perception: Vol. 7. Language and speech* (pp. 389–428). San Diego: Academic Press.

Gould, S. J. (2000). More things in heaven and earth. In H. Rose & S. Rose (Eds.), *Arguments against evolutionary psychology* (pp. 101–126). New York: Harmony Books.

Grafton, S. T., Mazziotta, J. C., Presty, S., Friston, K. J., Frackowiak, R. S., & Phelps, M. E. (1992). Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. *Journal of Neuroscience, 12(7),* 2542–2548.

Grafton, S. T., Woods, R. P., & Tyszka, M. (1994). Functional imaging of procedural motor learning: Relating cerebral blood flow with individual subject performance. *Human Brain Mapping, 1(3),* 221–234.

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature, 338,* 334–337.

Graybiel, A. M., Aosaki, T., Flaherty, A. W., & Kimura, M. (1994). The basal ganglia and adaptive motor control. *Science, 265,* 1826–1831.

Green, C. D., & Vervaeke, J. (1996). *What kind of explanation, if any, is a connectionist net?* North York, Ontaris, Canada: Captus Press.

Greenfield, P. M., & Savage-Rumbaugh, E. S. (1993). Comparing communicative competence in child and chimp: The pragmatics. *Journal of Child Language, 20,* 1–26.

Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Psychology, 6,* 270–292.

Greenwald, A. G. (1992). New look 3: Unconscious cognition reclaimed. *American Psychologist, 47,* 766–779.

Grimshaw, J. (1999). Lexicon. In *MIT encyclopedia of the cognitive sciences* (pp. 467-469). Cambridge, MA: MIT Press.

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science, 11(1),* 23–63.

Haack, S. (1979). Do we need fuzzy logic? *International Journal of Man-Machine Studies, 11,* 437–445.

Hartshorne, C., Weiss, P., and Burks, A. (Eds.). (1931–1958). *Collected papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Hayes, J. (1989). *The complete problem solver* (2nd ed.). Hillsdale, NJ: Erlbaum.

Hayes, P., & Ford, K. (1995). Turing test considered harmful. *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, 1,* 972–977.

Hebb, D. O., & McGill, U. (1949). *The organization of behavior: A neuropsychological theory.* Oxford, UK: Wiley.

Henderson, J. M., & Quandt, R. E. (1980). *Microeconomic theory: A mathematical approach* (3rd ed.). New York: McGraw-Hill.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science, 182,* 177–180.

Hinton, G. E. (1981). A parallel computation that assigns canonical object-based frames of reference. *Proceedings of the International Joint Conference on Artificial Intelligence.* Vancouver, BC, Canada.

Hockett, C. F. (1960). Logical considerations in the study of animal communication. In W. E. Lanyon & W. N. Tavolga (Eds.), *Animal sounds and communication.* Washington, DC: American Institute of Biological Sciences.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought.* Cambridge, MA: MIT Press.

Hopfield, J. J., & Tank, W. (1985). Neural computation of decisions in optimization problems. *Biological Cybernetics, 52,* 141–152.

Hosokawa, T., Rusakov, D. A., Bliss, T. V. P., & Fine, A. (1995). Repeated confocal imaging of individual dendritic spines in the living hippocampal slice: Evidence for changes in length and orientation associated with chemically induced LTP. *Journal of Neuroscience, 15,* 5560–5573.

Hubel, D. H. (1982). Exploration of the primary visual cortex, 1955–1978. *Nature, 299,* 515–524.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology, 160,* 106–154.

Hubel, D. H., Wiesel, T. N., & Stryker, M. P. (1978). Anatomical demonstration of orientation columns in macaque monkey. *Journal of Comparative Neurology, 177,* 361–379.

Hudson, P. T. W., & Postma, E. O. (1995). Choosing and using a neural net. In P. J. Braspenning, F. Thuijsman & A. J. M. M. Weijters (Eds.), *Artificial neural Networks: An introduction to ANN theory and practice* (pp. 273–287). Berlin: Springer-Verlag.

Humphreys, G. W., & Quinlan, P. T. (1987). Normal and pathological processes in visual object constancy. In G. W. Humphreys & M. J. Riddoch (Eds.), *Visual object processing: A cognitive neuropsychological approach* (pp. 43–105). Hillsdale, NJ: Erlbaum.

Humphreys, G. W., & Riddoch, M. J. (1984). Routes to object constancy: Implications from neurological impairments of object constancy. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 36A(3),* 385–415.

Humphreys, G. W., & Riddoch, M. J. (1987). *To see but not to see: A case study of visual agnosia.* Hillsdale, NJ: Erlbaum.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly, 32,* 127–136.

Jackson, P. (1999). *Introduction to expert systems* (3rd ed.). Boston: Addison-Wesley.

James, W. (1890). *The principles of psychology.* New York: Dover.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology, 21,* 60–99.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge, MA: Harvard University Press.

Johnston, W. A., & Heinz, S. P. (1978). Flexibility and capacity demands of attention. *Journal of Experimental Psychology: General, 107,* 420–435.

Jones, P. E. (1995). Contradictions and unanswered questions in the Genie case: A fresh look at the linguistic evidence. *Language & Communication, 15(3),* 261–280.

Jonides, J., & Smith, E. E. (1997). The architecture of working memory. In M. D. Rugg (Ed.), *Cognitive neuroscience* (pp. 243–272). Cambridge, MA: MIT Press.

Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., & Mintun, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature, 363,* 623–625.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Karmiloff-Smith, A. (2000). Why babies' brains are not Swiss army knives. In H. Rose & S. Rose (Eds.), *Arguments against evolutionary psychology* (pp. 173–188). New York: Harmony Books.

Katz, B. (1997). Annotating the world wide web using natural language. *Proceedings of the Fifth RIAO Conference on Computer Assisted Information Searching on the Internet.* McGill University, Montreal.

Kay, J., & Ellis, A. W. (1987). A cognitive neuropsychological case study of anomia: Implications for psychological models of word retrieval. *Brain, 11,* 613–629.

Kitzis, S. N. (2002). *Mind and meaning: A flight of imagination, a voyage of discovery.* Pacific Grove, CA: Wadsworth Thomson.

Klir, G. J., & Folger, T. A. (1988) *Fuzzy sets, uncertainty and information.* Upper Saddle River, NJ: Prentice Hall.

Kohler, W. (1920). *Static and stationary physical Gestalts.* Braunschweig, Germany: Vieweg.

Kohler, W. (1927). *The mentality of apes.* Berlin: Royal Academy of Sciences; New York: Harcourt Brace.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE, 78,* 1464–1480.

Kortge, C. (1990). Episodic memory in connectionist networks. *Proceedings of the Twelfth Annual Conference of the Cognitive Science Society,* 764–771, Hillsdale, NJ: LEA.

Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Psychology, 7,* 341–370.

Kosslyn, S. M. (1980). *Image and mind.* Cambridge, MA: Harvard University Press.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 47–60.

Kosslyn, S. M., & Shwartz, S. P. (1977). A simulation of visual imagery. *Cognitive Science, 1,* 265–295.

Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection.* Cambridge, MA: MIT Press.

Kremer, S. C. (1995). On the computational powers of Elman-style recurrent networks. *IEEE Transactions on Neural Networks, 6,* 1000–1004.

Kuczaj, S. A. (1978). Children's judgments of grammatical and ungrammatical past tense verbs. *Child Development, 49,* 319–326.

Kurzweil, R. (1999). *The age of spiritual machines: When computers exceed human intelligence.* New York: Viking Press.

LaBerge, D., & Buchsbaum, M. S. (1990). Positron emission tomographic measurements of pulvinar activity during an attention task. *Journal of Neuroscience, 10,* 613–619.

Lane, H. (1976). *The wild boy of Aveyron.* Cambridge, MA: Harvard University Press.

Lane, R. D., & Nadel, L. (2002). *Cognitive neuroscience of emotion.* New York: Oxford University Press.

Langton, C. (Ed.). (1989). *Artificial life.* Redwood City, CA: Addison-Wesley.

Lashley, K. (1950). In search of the engram. *Symposia of the Society of Experimental Biology, 4,* 454–482.

Launder, T. K. (1986). How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science, 10(4),* 477–493.

Levenson, R. W., Ekman, P., Heider, K., & Friesen, W. V. (1992). Emotion and autonomic nervous system activity in the Minangkabau of West Sumatra. *Journal of Personality and Social Psychology, 62,* 972-988.

Levine, D. N., & Calvanio, R. (1989). Prosopagnosia: A defect in visual configural processing. *Brain and Cognition, 10,* 149–170.

Levy, J., & Heller, W. (1992). Gender differences in human neuropsychological function. In A. A. Gerall, H. Moltz & I. L. Ward (Eds.), *Handbook of behavioral neurobiology.* New York: Plenum Press.

Lindsay, P. H., & Norman, D. A. (1972). *Human information processing: An introduction to psychology.* New York: Academic Press.

Llinas, R. R. (2002). *I of the vortex: From neurons to self.* Cambridge, MA: MIT Press.

Loftus, E. F. (1979). The malleability of human memory. *American Scientist, 67(3),* 312–320.

Loftus, E. F. (1995). Remembering dangerously. *Skeptical Inquirer, 19,* 20–29.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behaviour, 13,* 585–589.

Logie, R. H., Zucco, G., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica, 75,* 55–74.

Lorenz, K. (1994). *Man meets dog.* New York: Kondanska Globe.

Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1000–1014.

Lukasiewicz, J. (1967). Philosophical remarks on many-valued systems of propositional logic. In S. McCall (Ed.), *Polish Logic, 1920–1939.* Oxford, UK: Clarendon Press.

Maloney, J. C. (1999). Functionalism. In *MIT encyclopedia of the cognitive sciences* (pp. 332–334). Cambridge, MA: MIT Press.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57.*

Maren, A. J., Harston, C. T., & Pap, R. M. (1990). *Handbook of neural computing applications.* New York: Academic Press.

Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative Physiology and Psychology, 71,* 1–25.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London: Biological Sciences, 262(841),* 23–81.

Marr, D. (1982). *Vision.* San Francisco, CA : W. H. Freeman.

Martin, A., Wiggs, C. L., Ungerleider, L. G., & Haxby, J. V. (1996). Neural correlates of category-specific knowledge. *Nature, 379,* 649–652.

Marx, M., & Hillix, W. A. (1979). *Systems and theories in psychology* (3rd ed.). New York: McGraw-Hill.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18(1),* 1–86.

McClelland, J. L., McNaughton, B., & O'Reilly, R. C. (1994). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the

successes and failures of connectionist models of learning and memory. *Parallel Distributed Processing and Cognitive Neuroscience Technical Report,* PDP CNS.94.1.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? *Memory & Cognition, 6(4),* 462–472.

McCorduck, P. (1979) *Machines who think*. San Francisco: W. H. Freeman.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5,* 115–133.

McNamara, T. P. (1992). Priming and constraints in places on theories of memory and retrieval. *Psychological Review, 99(4),* 650–662.

McRae, K., & Hetherington, P. (1993). Catastrophic interference is eliminated in pretrained networks. *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society,* 723–728, Hillsdale, NJ: Erlbaum.

Mesulam, M.-M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology, 10,* 309–325.

Meyer, D. E, & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90,* 227–234.

Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., & Kramer, A. F. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research, 12,* 467–473.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in twenty-month-old infants. *Journal of Cognitive Neuroscience, 5(3),* 317–334.

Minsky, M. (1982). Why people think computers can't. *AI Magazine, (3)4,* Fall.

Minsky, M. (1985). *The society of mind.* New York: Simon & Schuster.

Minsky, M., & Papert, S. (1969). *Perceptrons.* Cambridge, MA: MIT Press.

Mitchell, M. (1996). *An introduction to genetic algorithms.* Cambridge, MA: MIT Press.

Miura, I. T., Okamoto, Y., Kim, C. C., & Steere, M. (1993). First graders' cognitive representation of number and understanding of place value: Cross-national comparisons: France, Japan, Korea, Sweden, and the United States. *Journal of Educational Psychology, 85(1),* 24–30.

Mohr, J. P. (1976). Broca's area and Broca's aphasia. In H. Whitaker & H. Whitaker (Eds.), *Studies in neurolinguistics* (pp. 201–233). New York: Academic Press.

Moravec, H. (1999). *Robot: Mere machine to transcendent mind.* Oxford, UK: Oxford University Press.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11,* 56–60.

Morris, R. G., Ahmed, S., Syed, G. M., & Toone, B. K. (1993). Neural correlates of planning ability: Frontal lobe activation during the Tower of London test. *Neuropsychologia, 31,* 1367–1378.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review, 76,* 165–178.

Morton, J. (1979). Facilitation in word recognition: Experiments causing change in the logogen model. In P. A. Kolers, M. Wrolstead & H. Bouma (Eds.), *Processing of visible language* (Vol. 1). New York: Plenum Press.

Musallam, S., Corneil, B. D., Greger, B., Scherberger, H., Andersen, R. A. (2004, July 9). Cognitive signals for neural prosthetics. *Science,* 305.

Myer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology, 90(2),* 227–234.

Myers, D. G. (2001). *Psychology.* New York: Worth.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review, 83,* 435–450.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

Neisser, U. (1976). *Cognition and reality.* San Francisco: W.H. Freeman.

Newell, A. (1991). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, NJ: Prentice Hall.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75,* 522–536.

Norman, D. A., & Shallice, T. (1980). Attention to action: Willed and automatic control of behavior. *Center for Human Information Processing Technical Report,* No. 99.

Notman, M. T., & Nadelson, C. C. (1990). *Women and men: New perspectives on gender differences.* Washington, DC: American Psychiatric Press.

Nudo, R. J., Milliken, G. W., Jenkins, W. M., & Merzenich, M. M. (1996). Use-dependent alterations of movement representations in primary motor cortex of adult squirrel monkeys. *Journal of Neuroscience, 16(2),* 785–807.

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a trade-off. *Parallel Distributed Processing and Cognitive Neuroscience Technical Report,* PDP CNS.94.4.

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart & Winston.

Palmer, J. A., & Palmer, L. K. (2002). *Evolutionary psychology: The ultimate origins of human behavior.* Boston: Allyn and Bacon.

Parkin, A. J. (1996). *Explorations in cognitive neuropsychology.* Oxford, UK: Blackwell.

Patterson, F. G. (1978). The gesture of a gorilla: Language acquisition in another pongid. *Brain & Language, 5(1),* 72–97.

Perrett, D., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research, 47,* 329–342.

Perrett, D. I., & Oram, M. W. (1993). Neurophysiology of shape processing. *Image and Visual Computing, 11,* 317–333.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology, 58,* 193–198.

Peterson, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature, 331,* 585–589.

Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1993). Functional activation of the human frontal cortex during performance of verbal working memory tasks. *Proceedings of the National Academy of Science, USA 90,* 878–882.

Pinker, S. (1997). *How the mind works.* New York: Norton.

Poole, D., Mackworth, A., & Goegel, R. (1998). *Computational intelligence: A logical approach.* Oxford, UK: Oxford University Press.

Popper, K. R., & Eccles, J. C. (1981). *The self and its brain.* Berlin: Springer.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32,* 3–25.

Posner, M. I. (1992). Attention as a cognitive and neural system. *Current Directions in Psychological Science, 1,* 11–14.

Posner, M. I., Inhoff, A. W., Friderich, F. J., & Cohen, A. (1987). Isolating attentional systems: A cognitive anatomical analysis. *Psychobiology, 15,* 107–121.

Posner, M. I., & Peterson, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25–42.

Posner, M. I., Sandson, J., Dhawan, M., & Shulman, G. L. (1989). Is word recognition automatic? *Journal of Cognitive Neuroscience, 1,* 50–60.

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 669–682). New York: Academic Press.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *The Journal of Neuroscience, 4,* 1863–1874.

Premack, D. (1976). *Intelligence in ape and man.* Mahwah, NJ: Erlbaum.

Principe, J. C., Euliano, N. R., & Lefebvre, W. C. (2000). *Neural and adaptive systems.* New York: John Wiley.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in extrastriate cortex studied by functional MRI. *Journal of Neurophysiology, 74,* 1192-1199.

Purpura, K. P., & Schiff, N. D. (1997). The thalamic intralaminar nuclei: A role in visual awareness. *Neuroscientist, 3,* 8–15.

Pylyshyn, Z. W. (1973). What the mind's eye tells the mind's brain: A critique of mental imagery. *Psychological Bulletin, 80,* 1–24.

Pylyshyn, Z. W. (1981). The imagery debate: Analog media versus tacit knowledge. In N. Block (Ed.), *Imagery* (pp. 151–206). Cambridge, MA: MIT Press.

Quinlan, J. R. (1986) Induction of decision trees. *Machine Learning, 1,* 81–106.

Rafal, R. D., Posner, M. I., Friedman, J. H., Inhoff, A. W., & Bernstein, E. (1988). Orienting of visual attention in progressive supranuclear palsy. *Brain, 111,* 267–280.

Rand, A. (1963). *For the new intellectual.* New York: Signet.

Rapaport, W. J. Center for Cognitive Science. State University of New York at Buffalo. Some definitions of "artificial intelligence." From http://www.cs.buffalo.edu/~rapaport/definitions.of.ai.html

Ratcliff, G., & Newcombe, F. (1982). Object recognition: Some deductions from the clinical evidence. In A. W. Ellis (Ed.), *Normality and pathology in cognitive functions* (pp. 147–171). London: Academic Press.

Reason, J. T. (1979). Action not as planned. In G. Underwood & R. Stevens (Eds.), *Aspects of consciousness* (pp. 67–89). New York: Academic Press.

Reed, S. K. (2000). *Cognition* (5th ed.). Belmont, CA: Wadsworth Thomson.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81,* 274–280.

Reisberg, D. (2001). *Cognition: Exploring the science of mind* (2nd ed.). New York: Norton.

Reisberg, D., Culver, C., Heuer, F., & Fischman, D. (1986). Visual memory: When imagery vividness makes a difference. *Journal of Mental Imagery, 10,* 51–74.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior, 14,* 665–681.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior, 12,* 1–20.

Roediger, H. L. (1980). The effectiveness of four mnemonics in ordering recall. Journal of Experimental Psychology: *Human Learning and Memory, 6,* 558-567.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York: Academic.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of experimental Psychology: General, 104,* 192–233.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review, 65,* 386–408.

Rozin, P., & Fallon, A. E. (1987). A perspective on disgust. *Psychological Review, 94,* 23–41.

Rubin, E. (1915). *Visuell wahrgenommene figuren*. Gyldenalske Boghandel, Copenhagen.

Ruckmick, C. A. (1913). The use of the term "function" in English textbooks of psychology. *American Journal of Psychology, 24,* 99–123.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group (Eds.) (1986). *Parallel distributed processing: Volume 1. Foundations.* Cambridge, MA: MIT Press.

Ryle, G. (1949). *The concept of mind.* London: Hutchinson.

Sandstrom, N., Kaufman, J., & Huettel, S. A. (1998). Males and females use different distal cues in a virtual environment navigation task. *Cognitive Brain Research, 6,* 351–360.

Savage-Rumbaugh, E. S. (1986). *Ape language: From conditioned response to symbol.* New York: Columbia University Press.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S., & Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development, No. 233, 58,* 3–4.

Saygin, A. P., Cicekli, I., Akman V. (2000). Turing test: Fifty years later. *Minds and Machines, 10,* 463–518.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures.* Hillsdale, NJ: Erlbaum.

Schiller, P. H., Sandell, J. H., & Maunsell, J. H. R. (1987). The effect of frontal eye field and superior colliculus lesions on saccadic latencies in the rhesus monkey. *Journal of Neurophysiology, 57,* 1033–1049.

Schultz, D. P., & Schultz, S. E. (1987). *A history of modern psychology* (4th ed.). Orlando, FL: Harcourt Brace Jovanovich.

Searching for the soul of the new machines. (2003, November 12). *International Herald Tribune.*

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts.* Cambridge, UK: Cambridge University Press.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3,* 417–457.

Searle, J. R. (1984). *Minds, brains, and science.* Cambridge, MA: Harvard University Press.

Searle, J. R. (1992). *The rediscovery of the mind.* Cambridge, MA: MIT Press.

Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception.* Indianapolis, IN: Bobbs-Merrill.

Sejnowski, T., & Rosenberg, C. (1986). *NETtalk: A parallel network that learns to read aloud* (Technical Report JHU/EEC-86/01). Baltimore: John Hopkins University Press.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In *The mechanisation of thought processes.* London: H. M. Stationery Office.

Shallice, T. (1982). Specific impairment of planning. *Philosophical Transactions of the Royal Society of London B, 298,* 199–209.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 153,* 652–654.

Sherry, D., Forbes, M., Khurgel, M., & Ivy, G. (1993). Greater hippocampal size in females of the brood parasitic brown-headed cowbird. *Proceedings of the National Academy of Sciences, U.S.A. 90,* 7839–7843.

Silveira, T. (1971). *Incubation: The effect of interruption timing and length on problem solution and quality of problem processing.* Unpublished doctoral dissertation, University of Oregon.

Silverman, G. (2001). Artificial intelligence in rehabilitation medicine: An emerging technology. *Proceedings of the IASTED International Conference.* Marbella, Spain.

Silverman, G. (2005, September). Human cognition and perceptual motor responses: An integration of iodiothetic and alothetic processes. *Proceedings of Biomedicine 2005.* Bologna, Italy.

Silverman, G., & Dworkin, B. R. (1991). Instrumental behavior shaping automata. In R. Weitkunat (Ed.), *Digital biosignal processing.* St. Louis, MO: Elsevier.

Silverman, I., & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind.* New York: Oxford University Press.

Silverman, I., & Phillips, K. (1998). The evolutionary psychology of spatial sex differences. In C. Crawford & D. L. Krebs (Eds.), *Handbook of evolutionary psychology* (pp. 595–612). Mahwah, NJ: Erlbaum.

Singer, W. (1996). Neuronal synchronization: A solution to the binding problem? In R. Llinas & P. S. Churchland (Eds.), *The mind-brain continuum: Sensory processes* (pp. 100–130). Cambridge, MA: MIT Press.

Singer, W. (1999). Binding by neural synchrony. In *MIT encyclopedia of the cognitive sciences* (pp. 81–84). Cambridge, MA: MIT Press.

Singer, W., Engel, A. K., Kreiter, A. K., Munk, M. H. J., Neuenschwander, S., & Roelfsema, P. R. (1997). Neuronal assemblies: Necessity, signature and detectability. *Trends in Cognitive Sciences, 1(7),* 252–261.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience, 18,* 555–586.

Skinner, B. F. (1957). *Verbal behavior.* New York: Appleton-Century-Crofts.

Smith, E. E. (2000). Neural bases of human working memory. *Current Directions in Psychological Science, 9,* 45–49.

Smith, E. E., & Jonides, J. (1994). Working memory in humans: Neuropsychological evidence. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1009–1020). Cambridge, MA: MIT Press.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283,* 1657–1661.

Smith, E. E., Jonides, J., Doeppe, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience, 7,* 337–356.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6,* 11–20.

Sowa, J. F. (1984). Conceptual Graphs. In *Conceptual structures: Information processing in mind and machine* (Addison-Wesley Systems Programming Series). Reading, MA: Addison-Wesley.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs, 74* (Whole No. 48).

Sperling, G. (1963). A model for visual memory tasks. *Human Factors, 5,* 9–31.

Sperry, R. W. (1985). The dual brain: Hemispheric specialization in humans. In D. F. Benson & E. Zaidel (Eds.), *UCLA forum in medical sciences,* No. 26 (pp. 11–26). New York: Guilford Press.

Steiner, J. E. (1979). Human facial expressions in response to taste and smell stimulation. *Advances in Child Development and Behavior, 13,* 257–295.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist, 57,* 421–457.

Stillings, N. A., Garfield, M. H., Rissland, J. L., Edwina, L., & Rosenbaum, D.A. (1995). *Cognitive science: An introduction.* Cambridge: M.I.T. Press.

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes.* New York: Raven Press.

Swinney, D., Zurif, E. B., Prather, P., & Love, T. (1996). Neurological distribution of processing resources underlying language comprehension. *Journal of Cognitive Neuroscience 8(2),* 174–184.

Tang, Y.-P., Shimizu, E., Dube, G. R., Rampon, C., Kerchner, G. A., Zhuo, M., et al. (1996). Genetic enhancement of learning and memory in mice. *Nature, 401,* 63–69.

Terrace, H. S., Petitto, L. A., Sanders, R. J., & Bever, T. G. (1979). Can an ape create a sentence? *Science, 206(4421),* 891–902.

Thagard, P. (2000). *Mind: Introduction to cognitive science.* Cambridge, MA: MIT Press.

Tinbergen, N. (1972). *Animal in its world: Explorations of an ethologist, 1932–1972.* Cambridge, MA: Harvard University Press.

Titchener, E. B. (1896). *An outline of psychology.* New York: Macmillan.

Tocco, G., Maren, S., Shors, T. J., Baudry, M., & Thompson, R. F. (1992). Long-term potentiation is associated with increased [3H]AMPA binding in rat hippocampus. *Brain Research, 573,* 228–234.

Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior, 3,* 449–459.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Treves, A., & Rolls, E. T. (1994). Computational analysis of the role of the hippocampus in memory. *Hippocampus, 4(3),* 374–391.

Triandis, H. C. (1994). *Culture and social behavior.* New York: McGraw-Hill.

Tsien, J. Z., Huerta, P. T., & Tonegawa, S. (1996). The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell, 87,* 1327–1338.

Turing, A. M., (1950). Computing machinery and intelligence, *Mind, 59, 236,* 433–460.

Turkle, S. (1995). *Life on the screen: Identity in the age of the internet.* New York: Touchstone .

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science, 185,* 1124–1131.

Tyler, L. K. (1992). *Spoken language comprehension: An experimental approach to disordered and normal processing.* Cambridge, MA: MIT Press.

Ungerleider, L.G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Vanier, M., & Caplan, D. (1990). CT-scan correlates of agrammatism. In L. Menn & L. Obler (Eds.), *Agrammatic aphasia: A cross-linguistic narrative sourcebook* (pp. 37–114). Amsterdam: John Benjamins.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, W. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science, 277,* 376–380.

Von der Malsburg, C. (1981). *The correlation theory of brain function.* Internal Report 81–2, Max-Planck-Institute for Biophysical Chemistry (Reprinted 1994).

Wachmuth, E., Oram, M. W., & Perrett, D. I. (1994). Recognition of objects and their component parts: Responses of single units in the temporal cortex of the macaque. *Cerebral Cortex, 4,* 509-522.

Wallas, E. (1926), *The Art of thought.* New York: Harcourt Brace.

Warren, R. M., & Warren, R. P. (1970). Auditory illusions and confusions. *Scientific American, 223(6),* 30–36.

Weizenbaum, J. (1976). *Computer power and human reason.* San Francisco: W.H. Freeman.

Weizenbaum, J. (1996). ELIZA—A computer program for the study of natural language communication between men and machines. *Communications of the ACM, 9,* 36–45.

Wellman, M. P. (1999). Multiagent systems. In *MIT encyclopedia of the cognitive sciences* (pp. 573–574). Cambridge, MA: MIT Press.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung, 4,* 301–350.

Whitby, B. (1996). The Turing test: AI's biggest blind alley? In P. Millican & A. Clarke (Eds.), *Machines and thought: The legacy of Alan Turing.* Oxford, UK: Oxford University Press.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings.* Cambridge, MA: Technology Press of MIT.

Wickens, D. D. (1972). Characteristics of word encoding. In W. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 191–215). New York: Winston.

Wiener, N. (1965). *Second edition: Or the control and communication in the animal and the machine.* Cambridge, MA: MIT Press.

Willshaw, D. J., & Buckingham, J. T. (1990). An assessment of Marr's theory of the hippocampus as a temporary memory store. *Philosophical Transactions of the Royal Society of London B, 329,* 205–215.

Wilson, E. O. (1999). *Consilience: The unity of knowledge.* New York: Random House.

Wilson, F. A., O'Scalaidhe, S. P., & Goldman-Rakie, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science, 260,* 1955–1958.

Wilson, S. W. (1985). Knowledge growth in an artificial animal. In *Proceedings of the first international conference on genetic algorithms and their applications* (pp. 16–23). Hillsdale, NJ: Erlbaum.

Windholz, G., & Lamal, P. A. (1985). Kohler's insight revisited. *Teaching of Psychology, 12,* 165–167.

Winston, P. H. (1993). *Artificial intelligence* (3rd ed.). Reading, MA: Addison-Wesley.

Witvliet, C. V. O., & Vrana, S. R. (1995). Psychophysiological responses as indices of affective dimensions. *Psychophysiology, 32,* 436–443.

Wojeiulik, E., & Kanwisher, N. (1999). The generality of parietal involvement in visual attention. *Neuron, 23,* 747–764.

Wood, G. (2003). *Edison's Eve: A magical history of the quest for mechanical life.* New York: Bantam-Doubleday-Dell Publishing Group.

Wood, G. (2002, February 16). Living dolls: A magical history of the quest for mechanical life. *The Guardian.*

Wood, N. L., & Cowan, J. (1995). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology: General, 124,* 243–262.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18,* 459–482.

Zadeh, L. (1965). Fuzzy sets. *Information and Control, 8(3),* 338–353.

Zeigarnik, B. (1938; original work published 1927). On finished and unfinished tasks. In W. D. Ellis (Ed.), *A source book of Gestalt psychology* (pp. 300–314). London: Routledge & Kegan Paul.

Zhang, S., & Schmitt, B. (1998). Language-dependent classification: The mental representation of classifiers in cognition, memory, and ad evaluations. *Journal of Experimental Psychology: Applied, 4(4),* 375–385.

Name Index

Abelson, R., 199, 366
Abrahamsen, A., 224
Ackley, D., 264
Adler, A., 82
Ahmed, S., 198
Akman, V., 341
Alivisatos, B., 194
Allen, W., 163
Alvarez, P., 192
Anderson, C. W., 224
Anderson, J. R., 135, 136, 150, 155, 230, 231, 249
Angell, J. R., 74, 76
Aosaki, T., 197
Arbib, M., 405
Archytas of Tarentum, 400
Aristotle, 31–32
Arkin, R., 427
Arkin, R. C., 420
Ashcraft, M. H., 96, 154
Atkinson, R. C., 134, 155
Au, T. K., 289
Averbach, E., 126
Awh, E., 194, 195

Babbage, C., 317, 318, 319
Baddeley, A., 128
Baddeley, A. D., 137, 138, 139, 194, 195
Bahrick, H. P., 132
Ball, T. M., 143
Banich, M. T. 185
Barr, A., 357
Barto, A. G., 224
Baudry, M., 189
Bavelier, D., 297
Bechtel, W., 224
Beer, R. D., 264
Bellugi, U., 99
Benson, D. F., 201
Berkeley, G., 33
Berlin, B., 289
Besner, D., 185
Biederman, I., 120, 121, 122
Bierce, A., 65
Bishop, D. V. M., 99
Black, M., 386
Blackmore, S., 443
Blackwell, B., 321
Block, N., 148, 312, 326, 344, 347, 348, 349
Bloom, A. H., 289
Bloom, P., 293
Boas, F., 449
Bochsbaum, M. S., 184
Boden, M., 59
Bond, A. H., 455
Bond, M. H., 448
Boster, J., 248
Bowlby, J., 241
Braitenberg, V., 405
Bransford, J. D., 201
Braspenning, P. J., 219
Breazeal, C., 401
Bringsjord, S., 344
Broadbent, D. E., 113, 114, 155

Broca, P., 294
Brooks, R., 404
Brooks, R. A., 323, 355, 400, 409, 416, 417, 425, 426, 444
Brown, J. A., 129
Brown, A. L., 201
Brown, A. S., 232
Bruce, C., 180
Bruner, J., 272
Buckingham, J. T., 192
Burks, A., 4
Burton, A. M., 180
Buss, D. M., 241, 244, 266, 441

Calvanio, R., 178
Campbell, D. T., 448
Campione, J. C., 201
Čapek, K., 401
Caplan, D., 298
Caramazza, A., 308
Carpenter, G. A., 221
Carr, H. A., 74, 75
Carroll, J. B., 288
Cawsey, A., 298
Chalmers, D., 49, 50
Chalmers, D. J., 323
Chapanis, A., 40
Chase, C. H., 221
Chase, W. G., 157
Cherry, E. C., 113
Chomsky, N., 96, 291, 292, 293, 325
Churchland, P., 11
Churchland, P. M., 54, 55, 56, 444
Cicekli, I., 341
Clark, V. P., 181.
Clark, A., 36
Clark, E. V., 276
Clark, H. H., 276
Coffey-Corina, S. A., 100
Cohen, A., 185
Colby, C. L., 184
Colby, K. M., 324
Collins, A. M., 226, 227, 228, 229, 230
Colmerauer, A., 337
Connelly, A., 196
Conrad, R., 130
Cook, V. J., 293
Cooper, L. A., 131
Copeland, J., 326
Corbett, G. G., 288, 289
Corkin, S., 190
Cosmides, L., 244, 251, 252, 267
Coull, J. T., 184
Cowan, J., 115
Coysevox, A., 314
Crick, F., 54, 444
Crowder, K. G., 126
Crystal, D., 293
Cummins, D. D., 253

Damasio, A., 55
Damasio, A. R., 308
D'Andrade, R., 248
Darwin, C., 19, 76, 239, 241, 242
Darwin, C. J., 126
Davies, I. R. L., 288, 289
Da Vinci, Leonardo, 313, 317
Dawson, M. R. W., 12, 223
Dayhoff, J., 219
De Jim, Strange, 95
Democritus, 33
Dennett, D. C., 57, 60, 61, 62, 148, 447
Deregowski, J. B., 448
Descartes, R., 34, 35, 60, 315, 316
Deutsch, D., 116
Deutsch, J. A., 116
De Vaucanson, 400
Devol, G., 400
Dewey, J., 74
Dhawan, M., 187
Dodson, J. D., 117
Donald, M., 270, 271
Donne, J., 446
Doupe, A. J., 285
Dover, G., 266
Dreyfus, H. L., 444
Dronkers, N. F., 297
Duhamel, J. L., 184
Dunbar, R., 256
Duncker, K., 24
Dworkin, B. R., 335, 409

Eals, M., 257
Eccles, J. C., 54

Edelman, G. M., 265
Edwards, F. A., 189
Ekman, P., 272
Ellis, A. W., 307
Elman, J., 223
Elman, J. L., 305
Ember, C. R., 449
Ember, M., 449
Emerson, R. W., 433
Engel, A. K., 202, 203
Engleberger, J., 400
Evans, A. C., 194

Fallon, A. E., 441
Farah, M. J., 140, 177, 180
Feinstein, M. H., 221
Fermat, ?, 345
Ferrara, R. A., 201
Fink, G. R., 184
Finke, R. A., 141
FitzGerald, R., 259
Flaherty, A. W., 197
Fodor, J. A., 99
Folger, T. A., 390
Forbus, K., 24
Ford, F., 346
Fox, J., 392
Frackowiak, R. S., 197
Freedman, D., 57
French, R. M., 223, 224
Freud, S., 62, 82, 83, 84
Friedrich, F. J., 185, 186
Frieze, I. H., 84
Friston, K. J., 197
Fromkin, V., 286, 287

Gabrieli, J. D. E., 169
Gadian, D. G., 196
Gainotti, G., 308
Gardner, B. T., 279
Gardner, H., 296, 449
Gardner, R. A., 279
Garfield, E. L., 221
Garfield, J. L., 29, 38, 39
Gasser, L., 455
Gaulin, S., 259
Gaulin, S. J. C., 247
Gelade, G., 108
Gentner, D., 24
Geschwind, N., 296
Gibson, J. J., 409, 445, 446
Gick, M. L., 26
Gigerenzer, G., 252, 254, 255
Glucksberg, S., 228
Gödel, K., 342
Goldberg, M. E., 184
Goldenberg, G., 140
Goldman-Rakic, P. S., 194
Goldstein, E.B., 181
Gould, S. J., 267
Grafton, S. T., 197
Gray, C. M., 203
Graybiel, A. M., 197
Green, C. D., 222
Greenfield, P. M., 280
Greeno, J. G., 154
Greenwald, A. G., 85
Grimshaw, J., 307
Grossberg, S., 221, 223
Gustavus Adolphus
(King of Sweden), 313

Haack, S., 390, 391, 392
Hartshorne, C., 4
Hawking, S., 318
Haxby, J. V., 308
Hayes, J., 149
Hayes, P., 346
Hebb, D. O., 188, 213, 214
Heinz, S. P., 116
Heller, W., 258
Henderson, J. M., 455
Herskivots, M. J., 448
Hetherington, P., 224
Hillix, W. A., 70
Hillyard, S. A., 116, 117
Hink, R. F., 116
Hinton, G. E., 180
Hitch, G., 139
Hockett, C. F., 256
Hoffrage, U., 254, 255
Holyoak, K. J., 24, 26
Hopfield, J. J., 220
Horney, K., 82
Hubel, D. H., 104, 166, 203
Hudson, P. T. W., 219

Huettel, S. A., 273
Hug, K., 252
Hume, D., 41, 42
Humphreys, G. W., 180

Imhoff, A. W., 185
Itard, J.-M.-G., 285

Jackson, F., 51
Jackson, P., 385
James, W., 75, 76
Jenkins, W. M., 197
Jernigan, T. L., 99
Jerrett, D. T., 289
Jerrett, T., 289
Johnson, J., 285
Johnson-Laird, P. N., 443
Johnston, W. A., 116
Jones, P. E., 286
Jonides, J., 139, 193, 194, 195
Joquet-Droz, P., 314, 316
Jung, C., 82

Kahneman, D., 117, 118, 119, 253, 254
Kanwisher, N., 181, 184
Karmiloff-Smith, A., 268, 269
Katz, B., 355
Kaufan, J., 273
Kay, J., 307
Kay, P., 289
Kempler, D., 179
Kimura, M., 197
Kitzis, S. N., 31
Klir, G. J., 390
Koch, C., 11, 54, 444
Koeppe, R. A., 193, 194, 195
Koffka, K., 77
Kohler, W., 77, 78, 80, 82
Kohonen, T., 220Konig, P., 203
Kortge, C., 224
Kosslyn, S. M., 5, 140, 141, 142, 143, 144, 147, 148
Kowalski, R., 337
Koza, J. R., 262
Kremer, S. C., 223
Kuczaj, S. A., 284
Kuhl, P. K., 285

LaBerge, D., 184
Lane, H., 285
Lane, R. D., 441
Langton, C., 263
Lashley, K., 187, 188
Law, K., 24
Lebiere, C., 135
Lenat, D., 357, 358, 400
Levenson, R. W., 442
Levine, D. N., 178
Levi-Strauss, C., 449
Levy, J., 258
Levy-Bruhl, L., 449
Littman, M., 264
Llinas, R. R., 55
Locke, J., 46, 47, 48
Loftus, E. F., 61, 85, 233
Logie, R. H., 139
Lorenz, K., 407, 408, 409
Louis XII (King of France), 313
Louis XIV (King of France), 314
Louis XV (King of France), 315
Luck, S. J., 117
Ludy, C. A., 297
Lukasiewicz, J., 386

McBurney, D. H., 247
McCarthy, J., 320, 325, 337
McClelland, J. L., 192, 222, 305
McCloskey, M. E., 228
McCorduck, P., 353
McCulloch, W. S., 212
McNamara, T. P., 226
McNaughton, B., 192
McRae, K., 224
Maloney, J. C., 40
Marcus, G. F., 284
Maren, A. J., 219
Maren, S., 189
Marler, P., 285
Marr, D., 10, 105, 106, 107, 192, 452
Marshal, W., 311
Martin, A., 308
Marx, M., 70
Maunsell, J. H. R., 184
Mazziotta, J. C., 197
Mendeleev, D., 70

Merzenich, M. M., 197
Mesulam, M. M., 186
Metzler, J., 131, 144, 145, 146
Meyer, D. W., 225
Meyer, E., 194
Milham, M. P., 185
Miller, G. A., 129
Milliken, G. W., 197
Mills, D. L., 100
Minoshima, S., 194, 195
Minsky, M., 215, 323, 326, 360, 365, 447, 456
Mintun, M. A., 194
Mishkin, M., 176
Mishkin, W., 196
Mitchell, M., 261
Miura, I. T., 290
Mohr, J. P., 297
Monides, J., 195
Moravec, H., 323, 400
Moray, N., 115
Morris, R. G., 198
Morton, J., 305
Musallam, S., 426
Mussa-Ivaldi, S., 401
Myers, D. G., 85

Nadel, L., 441
Nadelson, C. C., 258
Nagel, T., 50
Neisser, U., 95, 97, 409
Neville, H. J., 100
Newell, A., 153, 154, 155, 332, 337
Newport, E., 285
Newton, I., 318
Nilsson, N. J., 320
Nisbett, R., 253
Nobre, A. C., 184
Norman, D. A., 103, 116, 199, 200
Notman, M. T., 258
Nudo, R. J., 197

Offenbach, J., 316
Oram, M. W., 180
Oram, M. W., 122
O'Reilly, R. C., 192
O'Scalaidhe, S. P., 194

Paivio, A., 7, 8, 149
Palmer, J. A., 247, 256
Palmer, J. C., 61
Palmer, L. K., 247, 256
Papert, S., 215, 323
Parkin, A. J., 180.
Pascal, B., 317
Patterson, F. G., 279
Pavlov, I., 86, 87
Perret, D. I., 180
Perrett, D. I., 122
Peterson, L. R., 128, 129
Peterson, M. J., 128, 129
Peterson, S. E., 181, 298
Petrides, M., 194
Phelps, M. E., 197
Phillips, K., 257, 259
Picton, T. W., 116
Pinker, S., 255, 426
Pitts, W., 212
Plato, 31, 32, 46, 353
Poe, E. A., 275
Poole, D., 364
Popper, K. R., 54
Posner, M. I., 116, 181, 185, 186, 187
Potsma, E. O., 219
Premack, D., 279
Presty, S., 197
Principe, J. C., 357
Puce, A, 181.
Purpura, K. P., 54
Pylyshyn, Z. W., 8, 146, 147

Quandt, R. E., 455
Quillian, M. R., 226, 227, 228, 229, 230
Quinlan, J. R., 355, 393

Rafal, R. D., 186
Rand, A., 43, 44
Reason, J. T., 200
Redfern, B. B., 297
Reed, S. K., 114, 119
Reicher, G. M., 105
Reisberg, D., 142, 232
Reiser, B. J., 143
Richardson, ?, 349 (not in refs)

Riddoch, M. J., 180
Rips, L. J., 247, 248
Rissland, E. L., 221
Robinson, A., 337
Roediger, H. L., 49
Rolls, E. T., 192, 453
Rosch, E., 247
Rosenberg, C., 234, 235
Rosenblatt, F., 214
Ross, L., 253
Rozin, P., 441
Rubin, E., 93
Ruckmick, C. A., 76
Rumelhart, D. E., 222
Russell, B. 337
Russell, B., 1
Ryle, G., 36

Sacks, O., 176, 177
Sandel, J. H., 184
Sandson, J., 187
Sandstrom, N., 273
Sapir, E., 288
Savage-Rumbaugh, E. S., 280
Saygin, A. P., 341
Schank, R., 199, 366
Schiller, P. H., 184
Schmitt, B., 290
Schooler, L. J., 249
Schultz, D. P., 68, 74, 76, 77
Schultz, S. E., 68, 74, 76, 77
Schumacher, E. H., 195
Schvaneveldt, R. W., 225
Schwartz, S. P., 140
Schwent, V. L., 116
Scott, A. O., 399
Searle, J. R., 52, 53, 57, 58, 59, 303, 308, 323
Segall, M. H., 448
Sejnowski, T., 234, 235
Sejnowski, T. J., 11
Selfridge, O. G., 103
Seneca the Younger, 207
Shallice, T., 198, 199, 200
Shannon, C., 325
Shaw, G. B., 336, 337
Shelton, J., 308
Shepard, R. N., 131, 144, 145, 146
Sherry, D., 261
Shiffrin, R. M., 134, 155
Shoben, E. J., 247
Shors, T. J., 189
Shulman, G. L., 187
Silveira, T., 89, 90
Silveri, M. C., 308
Silverman, G., 327, 396, 409
Silverman, I., 257, 259
Simon, H. A., 153, 154, 157, 332, 337
Singer, W., 202, 203
Skinner, B. F., 85, 86, 87, 96
Smith, E. E., 139, 193, 194, 195, 247
Snyder, C. R. R., 116
Sowa, J. F., 365
Sowden, P. T., 289
Sperber, D., 449
Sperling, G., 126, 127
Sperry, R. W., 173
Squire, L. R., 192
Steiner, J. E., 47
Sternberg, S., 159
Stillings, N. A., 221
Stolz, J. A., 185
Stryker, M. P., 203
Stuss, D. T., 201
Sutton, R. S., 224
Syed, G. M., 198

Tang, Y.-P., 49
Tank, W., 220
Teller, A., 353
Terrace, H. S., 281
Tesla, N., 400
Thagard, P., 4, 22, 24, 440, 443, 450
Thompson, R. F., 189
Thorndike, E. L., 86
Tinbergen, N., 407, 408, 409
Titchener, E. B., 72, 73
Tocco, G., 189
Tolman, E. C., 88
Tooby, J., 244, 251, 252, 267
Toone, B. K., 198
Tranel, D., 308
Treisman, A., 108, 110, 111, 113, 115, 120
Treves, A., 192, 453
Triandis, H. C., 448

Triantafyllou, M., 400
Tsien, J. Z., 49
Turing, A., 311, 312
Turing, A. M., 337–346, 351
Turkle, S., 404, 405
Turvey, M. T., 126
Tversky, A., 253, 254
Twain, M., 125
Tyler, L. K., 99
Tyszka, M., 197

Ungerleider, L. G., 176, 308

Van Cantfort, T. E., 279
Vanier, M., 298
Van Paesschen, W., 196
Vargha-Khadem, F., 196
Vaucanson, Jacques de, 314
Vervaeke, J., 222
Victoria (Queen of England), 314
Voltaire, 314
Von der Malsburg, C., 203
Vrana, S. R., 442

Walker, J. A., 186
Wallas, E., 89
Wang, P. P., 99
Warren, R. M., 300
Warren, R. P., 300
Watkins, K. E., 196
Watson, J. B., 86, 406
Weiner, N., 430
Weisler, S. E., 221
Weiss, P., 4
Weizenbaum, J., 323, 345
Wellman, M. P., 455
Wernicke, C., 295
Wertheimer, M., 77, 78, 80
Whitby, B., 346
White, S., 326
Whitehead, A. N., 337
Whorf, B. L., 288
Wickens, D. D., 130
Wiener, N., 402
Wiesel, T. N., 104, 166, 203
Wiggs, C. L., 308
Willshaw, D. J., 192
Wilson, F. A., 194
Wilson, S. W., 263, 264
Winograd, T., 358
Witvliet, C. V. O., 442
Wojciulik, E., 184
Wood, N. L., 115
Woods, R. P., 197
Wundt, W., 69, 70, 71, 72, 73

Yerkes, R. M., 117

Zadeh, L., 386
Zeigarnik, B., 91
Zhang, S., 290
Zucco, G., 139

Subject Index

Abacus, 317
Abductive logic, 376
Acoustic code, 130
Acquisition, language of, 282–284
Act, 421–422
ACT*. *See* Adaptive Control of Thought (ACT*)
ACT. *See* Atomic components of thought (ACT)
Action-oriented perception, 411
Action potential, 170
Action schemas, 199
Activation function, 212–213
Activities of daily living (ADL), 322
Adaptive Control of Thought (ACT*), 135–137, 230–231
Adaptive Resonance Theory network (ART), 221
ADL. *See* Activities of daily living (ADL)
Affordance, 409–411
Agent link, 230
Agents, 447
"Agrammatic speech," 294
Agraphia, 286
AI. *See* Artificial intelligence (AI)
Alexia, 296
Algorithm, 10–11, 56
Algorithmic level, 11
Alphasic patient, 99
Altered mental states, and marijuana, 36
American Sign Language (ASL), 270, 277, 279
Amnesia, 190
Amygdala, 189
Analog, 3
Analogical reasoning, 24–26
Analog representations, 3, 7
Analog visual image, 4–5
Analogy, 24
Analytical Engine, 318–319
Android, 323
Angular gyrus, 296
Animat, 263–264
ANN. *See* Artificial neural network (ANN)
Anomia, 307
Anterior, 172–173
Anterograde amnesia, 190
Anthropology, 449
Aphasias, of language deficits, 294
Aphasic, 307
Apperceptive agnosia, 177–178
Applied artificial intelligence (AI), 323
Appropriate causal relation, 5
Arachnophobia. *See* Evolved psychological mechanism
Arbitrary, 276
Arcs, 364
Arcuate fasciculus, 295, 296
Arousal, 117
ART. *See* Adaptive Resonance Theory network (ART)
Articulatory loop, 137
Artificial intelligence (AI), xxi-xxii, 20–21, 150, 311
  agents, 47

applied, 323
artificial neural nets (ANNs), 392–393
"common sense," 358
computer as research tool for, 333–335
consciousness and, 56–59
defining, 320–330
definitional perspective, 311–352
design of intelligent agents, 356–357
distributed, 447
engineering goal of, 321
evaluating concept of, 322–330
Expert Systems (ESs), 380–386
general classes of, 321
goals of, 321
historical and philosophical roots, 312–320
intelligent agent, 408–410
linguistics and, 298–304
logical reasoning, 370–378
machine reasoning, 367–370
machine representation of knowledge, 357–367
methodologies, 331–332
operational perspective, 353–397
overall evaluation of, 345–346
practical world of, 354–356
programming language, 337
scientific method of, 321
strong, 57, 322–323, 408–410
Turing Test (TT) as operational definition of, 341, 344
weak, 57
working memory and, 438–439
Artificial life (A-Life), 239, 263–265
Artificial neural network (ANN), 19, 354, 392–393
back propagation and convergent dynamics, 215–219
characteristics of, 210–212
defined, 207
early conceptions of, 212–215
NETtalk, 234–236
principles underlying, 208–221
typologies, 219–221
See also Connectionist approach
Artificial retina, 214–215
ASL. *See* American Sign Language (ASL)
Assembly language program, 335
Assertions, in knowledge representation (KR), 363
Assertives, 303
Associationism, 48
Associative agnosia, 177–178, 180
Atomic components of thought (ACT), 24
Attention, xvi, 112–116, 436
attentuation model, 115–116
capacity model of, 117–119
classes of theories attempt to explain, 112–113
control of, 183
defined, 112
filter model, 113–115
information processing models of selection, 114
models of, 185–187
multimode model of, 116–117
neuroscience of, 181–187
Attention and Effort (Kahneman), 117
Attentuation theory, 115–116
AuRA. See Autonomous Robot Architecture (AuRA)
Automata, 312, 327
Automatic attentional processes, 199, 201
Automats, 322
Autonomous Robot Architecture (AuRA):
evaluation of, 429
schema, 427–428
Autonomous robotic explorers, 402
Axon, 170–171

Babbling stage, 283
Back propagation, 215–220
Back-tracking, 381
Backward Chaining, 376–378
Basal ganglia, 197
Base-rate fallacy, 253
Basic function, 212
Behavior:
in robotics, 406
learning in context of, 327
stimulus-response model of, 86
Behavior-based approach, to problem solving, 209

Behaviorism, 43, 85–88
  cognitive psychology *vs.,* 96–97
  evaluating, 87–88
  Turing Test (TT) and, 344, 347–349
Behaviorist, 66
Behaviorist theory of language, 96
Beliefs, infinite number of, 39
Binding, and neural synchrony, 202–203
Binding problem, 202
Biological plausibility, 221
Block approach, 347–349
Block-world problem, 155–156
Bottleneck theories, 112–113
Bottom-up model, 331–332, 356–357
Brain:
  cerebral cortex, 80
  consciousness neurons, 54
  illus., 13
  location of memories in, 187
  memory system, 189
  mind *vs.,* 30
  parietal lobe, 55
  thalamus, 54–55
  Wernicke-Geschwind model, 295
  See also Neuroscience
Brain anatomy, 170–175
  cingulate cortex, 182, 184–185
  cortex, 172–174
  directions in, 172–173
  frontal lobe, 182
  medulla, 183
  parietal lobe, 182, 184
  pons, 183
  posterior parietal lobe, 182
  split brain, 174–175
  superior collicutus, 182, 184
  thalamus, 182–184
  visual agnosias, 176–181
Brain damage:
  patients of, 99
  techniques for study of, 164–166
Brain imaging techniques, 167
  computerized axial tomography (CAT), 167
  magnetic resonance imaging (MRI), 168–169
  positron emission tomography (PET), 167–168
Brain recording techniques, 166–169
Brain regions, illus., 13
"Brainstorming," 262
Breadth-first search, 381–382
Broca's aphasia, 294
Broca's area, 294
Bugs, 385
Bus, 333–334

Caenorhabditis elegans, 325
Capacity model of attention, 117–119
Capacity theories, 113
"Capture errors," 200
Cartesian theater, 60
Cartographer, 427
Cases, 362, 366
CAT. *See* Computerized axial tomography (CAT) scan
Catastrophic interference, 223
Caudate nucleus, 189
Causation:
  entity model of, 43
  Humean model of, 41–42
Cell assembly, 214
Cell groupings, 214
Central Processing Unit (CPU), 60, 333–334
Cerebral cortex, 80
Cerebral hemispheres, 173
Chain-link problem, 89–91
Chaotic dynamics, 219
Chatbot, 340
Cheater-detection, 251–253
Children, 256
Chinese room scenario, 57–59
Chunking, 130
Cingulate cortex, 182, 184–185
Classical conditioning, 87
Classical dualism, 34
Classifiers, 263–264
Classifying, 38
Class membership, 390–391
Clearness, 73
Closed world paradigm, 405
Closure, principle of, 78
CLS. *See* Concept Learning Systems (CLS)
Coarticulation, 300
Cocktail party effect, 115
Cognition (Reed), 119

Cognition, linguistics and, 291–294
Cognitive architecture, 155–156
Cognitive demon, 103
Cognitive economy, 230
Cognitive map, 88
Cognitive neuroscience, 163
Cognitive penetration, 147
Cognitive processes, evolution and, 247–257
Cognitive psychologist, 66
Cognitive psychology, 17–18
  behaviorism *vs.*, 96–97
  defined, 97
  imagery debate, 147–149
  language and, 17
  memory, 125–126
  memory selection model, 116–117
  model-building approach, 119–120
  modularity of mind, 98–100
  overall evaluation of, 158
  overview, xix
  problem solving, 149–158
  process models, 98
  recognition-by-components, 120–122
  rise of, 95–97
  theories of attention, 112–116
  theories of visions and pattern recognition, 100–111
  theory of pattern recognition, 120–122
  visual imagery, 139–147
  working memory and, 435–436
Cognitive revolution, 96
Cognitive science (CS), 14
  as interdisciplinary approach, 2–3
  benefits of, 433–434
  consciousness and, 443–444
  elephant as metaphor for, 14
  emotions and, 441–443
  evaluating theories of, 449–452
  future of, 454
  importance of language in, 256
  individual and cultural differences, 448–449
  lack of single unified theory, 439–440
  multiagent systems, 455–457
  physical examination and, 444–446
  representation and, 3
  robotics, 399–431
  role of integration in, 452–454
  social environments, 446–448
Cognitive simulation, and natural language communication, 323–324
Commissives, 303
"Common sense," 358
Communication, 257, 276
Communication Bus, 333–334
Compatibilism, 45
Complex ideas, 46–47
Component process model, 185–186
Compositional semantics, 301
Computation, 9–12
  classical and connectionist views of, 12
  defined, 3
  tri-level hypothesis, 10–12
Computational level, 10
Computational level of description, 39
Computational modeling, 17
Computational theory of vision, 105–108
Computer, 3
  as model of brain organization, 334–335
  digital, 332–333
  increased use of, 97
  See also Central Processing Unit (CPU)
Computerized axial tomography (CAT) scan, 18, 97, 167
Computer language, 298, 335–336
Computer programming, 335–336
Computing Machinery and Intelligence (Turing), 311
Concept, 3, 22, 209, 363
Concept Learning Systems (CLS), 393
Conceptually-driven processes, 105
Conclusions, 377
Conditional statement, 23
Conditioned response (CR), 87
Conditioned stimulus (CS), 87
Conduction aphasia, 296
Conjunction fallacy, 253
Connectionism, defined, 207

Connectionist, 66
Connectionist approach:
  advantages of, 221–222
  evaluating, 221
  problems and disadvantages of, 222–224
Connectives, 367
Conscious, 82
Consciousness:
  artificial intelligence (AI) and, 56
  Cartesian theater explanation of, 60–61
  cognitive science (CS) and, 443–444
  defined, 49
  Freud's three-tiered system of, 82
  fundamental utilities of, 76
  mind as emergent property, 52
  mystery of, 49–52
  neurocomputational theory of, 55
  neurons, 54
  neuroscience and, 54–56
  nonmaterial, 53
  what-it's-like argument, 50–52
Consciousness Explained (Dennett), 60
Consilience: The Unity of Knowledge (Wilson), 439
Consolidation, 135, 190
Constancy problem, 120
Contention scheduling, 200–201
Contralateral side of brain, 174–175
Control group, 68
Controlled attentional processes, 199
Convergent dynamics, 215–219
Conversational pragmatics exercise, 308
Cooing stage, 282
Coronal plane, 173
Corpus callosum, 174
Cortex, 172–174, 196
Counterfactual statement, 289
CPU. *See* Central Processing Unit (CPU)
CR. *See* Conditioned response (CR)
Creative synthesis, 71
Critical period, 285
CS. *See* Cognitive science (CS)
CS. *See* Conditioned stimulus (CS)
Culture:
  cognitive science (CS) and, 448–449
  learning and, 268
  mythic, 271
  psychology and, 267–268
Cyc Project, 357–359

Data-driven processes, 104
Decay, 127
Decision demon, 13
Decision tree, 393, 395
Declarative knowledge, 24
Declarative memory, 131, 136, 196
Declaratives, 303
Deductive reasoning, 15
Deep representation, 142
Deep structure, 293
Defense mechanism, 84
"Degree of belief," 387–388
Degree of fan, 232
Deliberative behavior, 407
Demand characteristics, 147
Dendrites, 170–171
Dependent variable, 68
Deprivation, language, 284–288
Depth-first search, 381–382
Descartes approach, and mind/body debate, 316
Descriptionalist, 148
Determinism, 41–45, 84
Dichotic listening task, 113
Difference Engine, 318–319
Digital computer, 332–333
Digital representation, 6–7
Directives, 303
Displacement, 280
Dissociative identity disorder, 49
Distributed artificial intelligence, 447
Distributed coding, 181
Distributed cognition, 447
Distributed network models, 186–187
Distributed representation, 210
Dividend attention, 112
Divisible, 112
Doctrine of innate ideas, 47
Domain-general mechanisms, 48
Domain-specific mechanism, 48, 438
Dorsal, 172–173

Dorsal visual pathway, 176
Draw inferences, 367, 372
Dual-code hypothesis, 7–8
Dualism, 31–32
  critiques of, 36–38
  epiphenomenalist, 34–35
  interactionism, 35–36
  parallelism, 34
  technology and, 37
Duration, 73
Dynamic, 277
Dynamic system, 80

Early selection model, 115
EC. *See* Evolutionary computing (EC)
Echoic memory, 126
Echolocation, 50–51
Ecological models of cognition, 268–269
Ecological perception, 446
EEA. *See* Environment of Evolutionary Adaptation (EEA)
EEG. *See* Electroencephalogram (EEG)
Efficient computation, 365
Ego, 83–84
Electrical stimulation technique, 169
Electroencephalogram (EEG), 166–167
Elephant, as metaphor, 14
Embodiment, 444
Emergent property, 52–59
Emotions, cognitive science (CS) and, 441–443
Empiricism, 46
Emulator, 319
Encoding, 135
Enduring disposition, 118–119
Engineering goal of artificial intelligence (AI), 321
Engram, 187
Entity model of causation, 43
Environment:
  physical, 444–446
  social, 446–448
Environmental dependency syndrome, 197, 201
Environment of Evolutionary Adaptation (EEA), 241, 252
EP. *See* Evolutionary psychology (EP)
Epilepsy, 190
Epiphenomenalist, 34–35
Episodic memory, 132, 270
Epistemology, 30
Equipotentiality, 188
ERP. *See* Event-related potential (ERP)
Error signal, 215, 218
ESs. *See* Expert Systems (ESs)
Ethology, 407
Event-related potential (ERP), 116, 167
Evolution:
  stage theory of, 270–272
  theory of language, 256
Evolutionary computing (EC), 20, 239
  defined, 261
  methodologies, 261–263
Evolutionary psychology (EP), 19–20, 48, 240–241
  artificial life (A-Life), 263–265
  categorization, 247–248
  cognitive processes and, 247–257
  defined, 239
  evaluating, 266–269
  evolutionary computing, 261–263
  evolved psychological mechanisms, 244–246
  judgment under uncertainty, 253–257
  language, 256–257
  logical reasoning, 250–253
  memory, 248–250
  natural selection and genetics, 241–244
  Neural Darwinism, 265–266
  overall evaluation of, 269–270
  overview, xx
  sex differences in cognition, 257–261
  working memory and, 438
Evolved psychological mechanism, 244–246
Exaptation, 266
Executive control system, 137
Executive dysfunction, 197
Executive function, 197–201
  defined, 197
  theories of, 199–201
Exhaustive search, 160
Experiment, 144
Experimental group, 68

Expert Systems (ESs), 321, 354, 380–386
  breadth-first searching, 381–382
  depth-first searching, 381–382
  evaluation of, 384–386
  Fuzzy Logic, 386–392
  in action, 380–383
  MYCIN system, 383–384
  organization of, 381
Explanatory gap, 52
Explicit response, 86
Expressives, 303
External symbolic storage system, 271–272
Eye, illus., 101

Face perception, 180–181
Facial expressions, 273
Fallacies, 253–255
Feature demon, 103
Feature detection theory, 103–105
Feature integration theory, 108–111
Feed-forward network, 220
FFA. *See* Fusiform face area (FFA)
FIDO. *See* Field Integrated Design and Operations (FIDO)
Field Integrated Design and Operations (FIDO), 403
Field theory, 67, 80
Fissure, 174
Fixed-action patterns of response, 406
fMRI. *See* Functional magnetic resonance imaging (fMRI)
Focus, 44
Focused attention stage, 108
Formal logic, 356
Formal symbol manipulator, 12
Formula, 11, 363
Forward Chaining, 372–376
Frame, 362, 365–366
Free association, 236
Free will, 41–45
Freud, Sigmund:
  defense mechanisms, 84
  ego, 83
  iceberg model of mind, 83
  id, 82–83
  idealistic principle, 83
  pleasure principle, 83
  reality principle, 83
  repression, 84
  sublimation, 84
  superego, 83
  three-tiered system of consciousness, 82
Frontal lobe, 173–174, 182
Functionalism, 38–40, 74–77
  critics of, 40
  defined, 38
  evaluating, 76–77
  schools of thought in, 39
Functional kinds, 38
Functional magnetic resonance imaging (fMRI), 168–169, 181
Fundamental assumptions, 266
Fundamental utilities of consciousness, 76
Fusiform face area (FFA), 181
Fuzzy Logic, 321, 332, 354, 356, 386–392
  decision making or management, 390
  defined, 386
  "degree of belief" and, 387–388
  evolution of, 390–392
  knowledge-base (KB) and, 388
  methodology, 387
  rules, 389

Gambler's fallacy, 255
Gathering process, communication and, 257
Gender:
  exploring environment, 273
  spatial ability and, 259
  understanding facial expressions, 273
General and fortress problem, 25
Generalization, 222
Generalized delta rule, 215
General Problem Solver (GPS), 153–154, 415
General-purpose processor, 244
General-purpose solutions, 330
Generation, of language, 298
Generative, 277
Genetics, 241–244
Geons, 121–122

Gestalt psychology, 16
  evaluating, 81–82
  field theory, 67, 80
  focus of, 66, 77–82
Gestalt theory, 435
GPS. *See* General Problem Solver (GPS)
Graceful degradation, 222
Grammar, 278
  phrase structure, 291–292
  role of, in linguistics, 291–294
  transformational, 292
  universal, 293
Guided search, 233
Gyrus, 174

"Hasa" link, 230
Hebb rule, 188–189, 213
Heuristic, 151–152, 253
Hidden layer, 215, 217–218, 220
Hierarchical isa structure, 361
Hierarchical organization, 226–227
Hierarchical paradigm, 414–416, 427
Hierarchical semantic network, 226–230
High Level Languages (HLL), 335–336
Hippocampal system, 189–193
History of Modern Psychology, A (Schultz and Schultz), 68
HLL. *See* High Level Languages (HLL)
Homeostatic Control, 428
Homo sapiens, 240, 256
Homunculus problem, 61
Hopfield-Tank network, 220
Hopi language, 288
Horizontal plane, 173
Human behavior, 43. *See also* Behaviorism
Human emotions, five basic, 27
Humean model of causation, 41–42
Hybrid deliberative/reactive paradigm, 420–425
Hybrid model, 421–422
Hybrid paradigm, 420–421
Hybrot, 425
Hypothalamus, neural activity in, 223
Hypothesis, 7–9, 67–68. *See also entries for individual hypotheses*
Hypothetic-deductive approach, 67

Iceberg model of mind, 83
Iconic memory, 126
Id, 82–84
ID3 algorithm, 393–394
Idealism, 33
Idealistic principle, 83
IG. *See* Imitation game (IG)
Illumination stage, 89
Image demon, 103
Image generation, 142
Image inspection, 142
Image processes, 142–147
Imagery, xvi, 140. *See also* Visual imagery
Imagery debate, 147–149
Images, as processed spatially, 148
Image transformation, 144
Imitation game (IG), 336, 338–341, 350
Immediate experience, 70
Implementational level, 11
Implicit response, 86
Incompatibilism, 45
Incubation stage, 89–90
Independent variable, 68
Inductive reasoning, 15, 379
Infants, 47
Inferotemporal cortex (IT), 180
Information Processing Language (IPL), 336
Information processor, 3, 97–98
Inherent fragment of intelligent reasoning, 361
Inheritance, 22
Innate behavior, 47–48
Innate reasoning mechanism (IRM), 408–409
Input layer, 215–218, 220
Input/Output (I/O), 333–334
Insight learning, 78, 89–90
Integrated program of study, for working memory, 434–439
Integration, role of cognitive science (CS) in, 452–454
Intelligent agents, design of, 356–357
Intensity, 73
Intentionality, 4–5
Interactionism, 35–36

Interdisciplinary perspective:
artificial intelligence approach, 20–21
cognitive approach, 17–18
evolutionary approach, 12–21
linguistic approach, 20
network approach, 18
neuroscience approach, 18
philosophical approach, 15–16
psychological approach, 16–17
robotics approach, 21
See also entries for specific approaches
Interference, 130, 222
Internal perception, 70
Intervening variables, 88
Intraparietal sulcus, 184
Introspection, 70
IPL. *See* Information Processing Language (IPL)
Ipsilateral side of brain, 175
IRM. *See* Innate reasoning mechanism (IRM)
"Isa" link, 230
Isomorphism, 4, 80, 139–140
IT. *See* Inferotemporal cortex (IT)

Jet Propulsion Laboratory (JPL), 403
JLP. *See* Jet Propulsion Laboratory (JPL)
Judgment, xvi
Judgment under uncertainty, 253–257

KB. *See* Knowledge-based (KB) approach
Knowledge, machine representation of, 357–367
Knowledge acquisition problem, 45–49
Knowledge-based (KB) approach, 209, 357–358, 380
Knowledge representation (KR), 19, 354
assertions, 363
characteristics of, 360–362
concept, 363
frames, 365–366
in formulas, 363
objectives of, 359–360
predicates, 362–364
production rules, 364
technologies, 362
Kohonen network, 220
Kosslyn and Schwartz Theory, 146–147
KR. *See* Knowledge representation (KR)

Language, 99
as digital mental representation, 6
behaviorist theory of, 96
characteristics of, 276–277
cognitive psychology and, 17
computer, 335–336
evolutionary psychology and, 256–257
generation of, 298
grammar of, 278
importance of, 275–276
nature of, 276–278
shared representations, 447–448
symbolic elements of, 277
use of, in primates, 278
Language acquisition, 282–284
Language deficits, 36, 294
Language deprivation, 284–288
Language-deprived children, 287
Language evolution, theory of, 256
Latent learning, 88
Lateral, 172
Late selection model, 116
Lattices, 78–79
Learning:
basal ganglia role in, 197
culture and, 268
defined, 188
in context of behavior, 327
machine-based algorithm for skill, 327
memory and, 188–189
operant, 327
rote, 327
Left parietal cortex, 189
Lesion study, 165
Letters, as symbols of language, 6
Lexicon, 305
modern conceptions of, 307–308
morpheme within, 307
Lexigrams, 279–280
Libertarians, 45
Life on the Screen (Turkle), 404
Limbic cortex, 196

Linguistic approach:
importance of language, 20, 275–276
language use in primates, 278–281
overall evaluation of, 304
overview, xxi
working memory and, 438
See also Language; Linguistics
Linguistic relativity hypothesis, 288–291
Linguistics:
artificial intelligence (AI), 298–304
cognition and, 291–294
defined, 275
neuroscience and, 294–298
philosophies and, 288–291
Links, 210–211, 230
Literal encodings, 142
Lobes, 174. *See also specific types of lobes*
Local minimum, 219
Local representation, 210
Logic, 22–23
Logical reasoning, 250–253, 370–378
Backward Chaining, 376–378
draw inferences, 372–378
evaluation of rule-based reasoning, 379–380
Forward Chaining, 372–376
inductive reasoning, 379
sub-goal, 376–377
Logogen, 305–308
Logogen model of word recognition, 304–308
Long-term memory, 131–133, 155, 436
computer, 334
neural substrates of, 196–197
Long-term potentiation (LTP), 189
Loss function, 219
Louis XV, 315
LTP. *See* Long-term potentiation (LTP)

Machine, man as, 315–316
Machine reasoning, 367–370
Machine representation of knowledge, 357–367
"Magical number seven," 129
Magnetic resonance imaging (MRI), 18, 97, 168–169
Main memory, 334
Man as machine, 315–316
Man Who Mistook His Wife for a Hat, The (Sacks), 176
Marijuana, and altered mental states, 36
"Mark I," 214
Materialism, 33
Maximal onset principle, 293
Means-ends analysis, 151–152, 330
Mechanical calculators, 317
Mechanical computation, 316–320
Mechanical dolls, 313–314
Mechanism, 82–85
Medial, 172
Medial temporal lobe, 196
Mediate experience, 70
Medium of human expression, 362
Medulla, 183
Memory, xvi
ACT* model of, 135–137
brain areas that form part of, 189
computer, 333–334
declarative, 131, 136, 196
defined, 125
echoic, 126
episodic, 132, 270
evolutionary psychology and, 248–250
iconic, 126
learning and, 188–189
location in brain for, 187
long-term, 131–133, 196–197
modal model of, 134–135
models of, 133–139
neural substrates of long-term, 196–197
primary and recency effects exercise, 161
procedural, 131, 196
production, 136
reconstructive, 233
semantic, 131–132, 270
sensory, 126–127
SOAR model and long-term, 155
See also Working memory
Memory selection model, 116–117
"Mental force fields," 80
Mentality of Apes, The (Kohler), 78
Mental operations, study of, 76

Mental representation, categories of, 22–24
Mental rotation, 140
Message attenuation, 115
Metacognition, 201
Metacognitive regulation, 201
Meta knowledge, 321
Metaphor, elephant as, 14
Metaphysics, 30
Method of Difference, 318
Mimetic skill, 271
Mind:
  as black box, 85–88
  as emergent property, 52–59
  as information processor, 97–98
  as "Swiss army knife," 244
  brain *vs.*, 30
  Freud's iceberg model of, 83
  modularity of, 98–100
  operation of, 66
  theory of, xv
Mind (Thagard), 440
Mind: Introduction to Cognitive Science (Thagard), 22
Mind/body debate, 316
Mind-body problem, 30–34
Missionary-Cannibal problem, 154
Mission planner, 427
Modal memory model, 134–135
Model-building approach, 119–120
Model of brain organization, computer as, 334–335
Models of attention, 185–187
Models of mind, xv
Modular approach, and brain-damaged patients, 99–100
Modularity of mind, 17
Modularity of Mind, The (Fodor), 99
Molecular drive, 267
Molecules, illus., 13
Momentary intention, 119
Monism, 31, 33–34
Morphemes, 277, 307
Morphology, 278, 293
Mosaic model of development, 269
MRI. See Magnetic resonance imaging (MRI)
Multiagent system, 455–457
Multi-layer network, 220
Multimode model of attention, 116–117
Multiple Drafts Theory of Consciousness, 60–63
Multiple-unit recording, 166
MYCIN system, 383–384
Myth, 271
Mythic cultures, 271

Narrative thought, 272
NASA, 402–403
Nativism, 46
Natural language, 298–299
Natural language communication, cognitive simulation and, 323–324
Natural language processing, 298
  evaluation of, 303–304
  speech recognition, 299–303
Natural selection, 19, 241–244
Nature-nurture debate, 49
Navigator, 427
Neocortex, and language evolution, 256
Neo-Freudian, 82, 84
Nervous system, structural levels of analysis in, 13
NETtalk, 234–236
Network approach, 18–19, 207–237
  artificial neural network (ANN), 207–221
  overall evaluation of, 233–234
  overview, xx
  working memory and, 447
  See also Connectionist approach
Neural Darwinism, 20, 239, 265–266
Neural functions exercise, 203–204
Neural group selection, 265
Neural net, 356
Neural network, 321
  illus., 13
  simple, 211
  See also Artificial neural network (ANN)
Neural processing model. *See* Working memory
Neural substrates:
  of long-term memory, 196–197
  of spatial working memory, 194
  of working memory, 193–197

Neural synchrony, 202–203
Neurocomputational theory of consciousness, 55
Neuron, 2, 54
  anatomy of, 170–171
  illus., 13
Neuron anatomy, 170
Neuropsychology, 163
Neuroscience, 18, 437
  linguistics and, 294–298
  brain anatomy, 170–175
  consciousness and, 54–56
  defined, 163
  models of attention, 185–187
  neuron anatomy and physiology, 170
  of attention, 181–187
  of executive function and problem solving, 197–201
  of memory, 187–196
  of visual object recognition, 175–181
  overall evaluation of, 201–202
  overview, xix
  perspective, 163–164
  semantic network, 224–223
Neuroscience methodology, 164–189
  brain recording techniques, 166–169
  electrical stimulation technique, 169
  techniques for study of brain damage, 164–166
Neurotransmitters, 170
Neutral drift, 266
Node, 210–212, 220, 231–232, 364
Nonfluent aphasia, 294
Nonmaterial consciousness, 53
Nonreductive physicalist, 33

Object-centered description, 106
Object constancy problem in perception, 106
Object link, 230–231
Object recognition, xvi, 436
Object representations, 106–107
Occam's razor, 36
Occipital lobe, 173–174
One-word stage, 283
On the Origins of Species (Darwin), 241
Ontology, 361
Open world paradigm, 405
Operant conditioning, 87, 327
Operating system, of personal computer (PC), 335
Operational perspective, 394–396
Operator, 150
Ordinate category, 227
Origins of the Modern Mind (Donald), 270
Oscillatory dynamics, 219, 223
Output layer, 215–218, 220

Pandemonium model, 103
Paradigm, 321
Paradigmatic thought, 272
Parallel distributed processing, 208
Parallelism, 34
Parallel memory search, 159
Parallel search, 108
Parietal lobe, 55, 174, 184
Parry, 324
Partial-report condition, 127
Partial results, 334
Pattern recognition, 100, 120–122, 209
PC. *See* Personal computer (PC)
Perception, 100, 330
  object constancy problem in, 106
  representational-computational *vs.* sensory-motor view of, 445
Perception and Communication (Broadbent), 113
Perceptrons, 214–216, 220
Perceptual categorization deficit, 178
Perceptual organization, principles of, 78–79, 81
Perceptual pop-out, 108–109
Personal computer (PC), 318
Pessimists, 45
PET. *See* Positron emission tomography (PET) scan
Phase sequence, 214
Phenomenal concept of mind, 50
Phenomenology, 77
Philosophical approach, 15–16, 29–64
  as non-empirical approach, 59–60
  dualism, 34–38
  free will-determinism debate, 41–45
  functionalism, 38–40

knowledge acquisition problem, 45–49
mind as emergent property, 52–59
mind-body problem, 30–34
monism, 33–34
Multiple Drafts Theory of Consciousness, 60–63
mystery of consciousness, 49–50
overall evaluation of, 59–60
overview, xviii
working memory and, 435
Philosophy, 15
defined, 29
linguistics and, 288–291
Phoneme, 277, 293, 299
Phonology, 278, 293
Phrase structure, 291–292
Phrase structure grammar, 291–292
Physical environments, cognitive science (CS) and, 444–446
Physicalism, 33–34
Physical kind, 38
Physical Symbol System Hypothesis, 332
Physiology, 170
Pictorialist, 148
Pineal gland, 35
Plan, 421–422, 427
Pleasure principle, 83
Pons, 183
Pop-out, 108–109
Positron emission tomography (PET) scan, 18, 97, 167–168, 193
Posterior, 172–173
Posterior parietal lobe, 182
"Post-natal neural assembly," 269
Pragmatic analysis, 299, 302–303
Pragmatics, 303
Pragnanz, 78
Preattentive stage, 108
Preconscious mind, 82
Predicate calculus, 8
Predicate language, diagram of, 368
Predicates, 362–364
Prediction, 42
Preparation stage, 89
Primary effect, 161
Primary motor cortex, 173–174
Primary somatosensory cortex, 173–174
Primates:
American Sign Language (ASL) and, 270, 279
language use in, 278–282
Priming, 225
Principia Mathematica (Russell, Whitehead), 337
Principle of closure, 78
Principle of proximity, 78, 81
Principle of similarity, 78, 81
Principles of perceptual organization, 78–79, 81
Proactive interference, 130
Problem solving, xvi, 149–158, 330, 436
artificial intelligence (AI), 150
behavior-based approach, 209
characteristics of, 150
defined, 149
general problem solver model (GPS), 153–154
heuristic, 151–152
neuroscience of executive function and, 197–201
SOAR model, 155
Problem space, 150–151
Procedural knowledge, 24
Procedural memory, 131, 196
Procedures, 365
Process model, 98
Production memory, 136
Production rule, 23, 364
Program language, artificial intelligence (AI), 337
Programming. *See* Computer programming
Proposition, 3, 22–23, 370
Propositional encodings, 142
Propositional hypothesis, 8
Propositional representations, 8–9
Prosopagnosia, 180–181
Prototype, 229
Proximity, principle of, 78, 81
Psychoanalysis, 67
Psychoanalytic psychology, 16, 82
anxiety, 85
evaluating, 84
Psychological approach, xviii, 65–93
evaluation of, 88–89

functionalism, 74–77
Gestalt movement, 71–72
insight learning, 89–91
structuralism, 72–74
to working memory, 435
voluntarist movement, 68–72
Psychological concept of mind, 50
Psychological disorder, major categories of, 272
Psychological inertia, 197
Psychological relations, 76
Psychologism, 347–348
Psychologist's fallacy, 75
Psychology:
behaviorism, 85–88
culture and, 267–268
defined, 65–67
mechanism and psychoanalytic, 82–85
mind as black box, 85–88
scientific method and, 65, 67–68
Punishment, 87
Putamen and globus pallidus, 189

Qualia, 40, 51
Quality, 73
Quasi-pictorial, 148

RAS. *See* Reticular activating system (RAS)
Rationalism, 46
Raw primal sketch, 106
rCBF. *See* Regional cerebral blood flow (rCBF)
Reactive paradigm (subsumption), 416
Reactive response, 406
Reagent, 73
Reality principle, 83
Reasoning, xvi, 329
Recency effect, 161
Reciprocal altruism, 252
Recognition-by-components, 120–122
Recognition of stimulus object, 101
Reconstructive memory, 233
Record, 366
Recurrent network, 220, 223
Rediscovery of Mind, The (Searle), 52
Reductionism, 33, 53
Reductive physicalism, 33
Referents, 4–5
Reflexes, 47–48
Reflexive responses, 406
Regional cerebral blood flow (rCBF), 198
Regulatory model of development, 269
Rehearsal, 128, 213
Reinforcement, 87
Relational graphs exercises, 429
Relation link, 231
Replication, 42
Representation, 3–9
as semantic, 4
as symbolic, 4–5
crucial aspects of, 4
defined, 359
digital, 6–7
dual-code hypothesis, 7–8
Representational-computational, *vs.* sensory-motor views of perception, 445
"Representation bearer," 4
Representativeness heuristic, 253
Repression, 84
Reticular activating system (RAS), 182–183
Retrieval, 135
Retrieval cues, 225
Retroactive interference, 130
Retrograde amnesia, 190
Reverse engineering, 267
Robot, 323, 399
Roboticist, 401, 410
Robotic paradigm, 412–414
hierarchical, 414–416
hybrid deliberative/reactive paradigm, 420–425
reactive paradigm, 416–420
Robotics, 21, 399–431
achievements in, 402–404
action-oriented perception, 411
Autonomous Robot Architecture (AuRA), 426–429
behavior and, 406
biological and behavioral foundations, 404–412
deliberative behavior, 407
evaluation of biological basis of, 411–412

fixed-action patterns of response, 406
highlights in history of, 400–401
overall evaluation of, 425–426
overview, xxii
primitive functions, 412–413
reactive responses, 406
reflexive responses, 406
transferring behavioral models to, 410–411
working memory and, 439
Rote learning, 327
Rule, 3, 23–24
Rule-based reasoning, 379–380
Russell, Bertrand, 1

Sagittal plane, 173
SAIL. *See* Stanford Artificial Intelligence Laboratory (SAIL)
SAS. *See* Supervisory Attentional System (SAS)
Schema, 199–200, 427–428
Scientific goal of artificial intelligence (AI), 321
Scientific method, 67–68
Scientific model, 16
Scripts, 200, 362, 366–367
Selection, 242
Selection theory, 19
Selective attention, 112
Self-terminating search, 159–160
Semantic, 4, 278
Semantic analysis, 301–302
Semantic code, 131
Semantic memory, 131, 270
Semantic model of data representation, 364–365
Semantic network, 19, 362, 364–365
characteristics of, 225–233
defined, 224
evaluating, 231–233
hierarchical, 226–230
propositional, 230–233
Sensations, attributes of, 73
Sense, 421–422
SENSE-ACT modules, 421
SENSE-ACT primitives, 419
SENSE-PLAN-ACT primitives, 412–413, 415–419
Sensory memory, 126–127
Sensory-motor views of perception, representational-computational *vs.*, 445
Sentence verification rule, 227
Sequencing, 197–198
Serial memory search, 159
Serial processors, 208–209
Serial search, 110
Sex differences, in cognition, 257–261
Sexual division of labor, 257
Sexual selection, 243
Shared representation, language, 447–448
Shiftability, 112
Shift transformation, 144
Short-term memory, 127, 130
Sigmoid activation function, 212–213
Similarity, principle of, 78, 81
Simple ideas, 46
Simple neural network, 211
Single-cell recording, 166
Single-layer network, 220
Skill learning, 197, 327
"Slave system," 138
Slot, 365
SOAR model, 24, 155–158
evaluating, 156–158
long-term memory and, 155
Society of Mind, The (Minsky), 447
Solipsism, 33
Spandrel, 267–268
Spatial ability, gender and, 259
Spatial working memory, 194
Special-purpose solution, 330
Specificity coding, 181
Speech recognition, 299–303
pragmatic analysis, 302–303
semantic analysis, 301–302
syntactic analysis, 301
Speech spectrogram, 299–300
Spinal reflex, neural connections in, 48
Split brain, 174–175
Spoken statements, types of, 303
Spreading activation, 225, 227–228, 233
Stability-plasticity dilemma, 223
Stage theory, 105
Stage theory of evolution, 270–272

Stanford Artificial Intelligence Laboratory (SAIL), 415
Static and Stationary Physical Gestalts (Kohler), 80
Stereoscopic vision, 269
Stimulus error, 72
Stimulus-response model of behavior, 86
Storage, 135
Stream of consciousness, 75
Striatum, 189
Strips, 415
Strong artificial intelligence (AI), 322–323, 408–410
Stroop effect, 184–185
Structuralism, 72–74
Structured, 276–277
Subgoal, 150, 376–377
Sublimation, 84
Subordinate category, 227–229
Substantive thought, 76
Subsumption, 416
Subsumptive Architecture, 418, 444–445
Sulcus, 174
Superego, 83
Superior colliculus, 182, 184
Superordinate category, 227–228
Supervised network, 219
Supervisory Attentional System (SAS), 200
Surface representation, 141
Surface structure, 293
Surrogate, 361
Sustainability, 112
"Swiss army knife," mind as, 244
Symbolic, 4–5
"Symbolic AI," 332
Symbols, 4–5, 276
  external symbolic storage system, 271–272
  letters for language, 6
  ordering of, 276–277
Synapse, 170–171
  cross-section of, 172
  illus., 13
Synaptic cleft, 170
Synaptic plasticity, 188
Synchrony hypothesis, 203
Syntactic analysis, 299, 301
Syntax, 6, 278, 293, 307, 370

Tabula rasa, 46
"Tales of Hoffman, The" (Offenbach), 316
Taxes, 406
Teacher, The, 214, 218, 220
Technology, and dualism, 37
Template matching theory, 100–102
Temporal lobe, 173–174
Temporal synchrony, 203
Terminal button, 170–171
Thalamus, 54–55, 182, 184
Theory, 68
Theory of mind, xv
Theory or natural selection, 241
3-D sketch, 106
Threshold, 115, 211
Threshold of excitation, 170
Throughput, 334
"Tip of the tongue," 231–232
Token (instance), 231
Top-down model, 331–332, 356
T.O.T. phenomenon, 231–232
Tower of Hanoi problem, 152–153
Tower of London problem, 197–198, 201
TRACE, 305
Transformational grammar, 292
Transformation rules, 293
Transitive thought, 76
"Treatise on Man," 315–316
Trial-and-error learning, 327
Tridimensional theory of feeling, 71
Tri-level hypothesis, 10–12, 452
Truth Table, 389
Truth Tables, 371
TT. Turing Test, 325
"Tumor problem," 24–25
Turing Test (TT), 325, 336–346
  arguments against, 342–343
  as operational definition of artificial intelligence (AI), 341, 344
  behaviorism and, 344, 347–349
  evaluation of, and Turing detractors, 341

future of, 341
sketch of, 339
too easy or too difficult, 344–345
2 1/2-D sketch, 106
Two-word stage, 283
Type, 231
Typicality effect, 247

Unconditioned response (UCR), 87
Unconditioned stimulus (UCS), 87
Unconscious mind, 82
Universal grammar, 293
Universal quantifiers, 367
Unsupervised network, 219

Variable, 367, 371
Variation, 241
Ventral, 172–173
Ventral visual pathway, 176
Verbal Behavior (Skinner), 96
Verification stage, 89
Viewer-centered description, 106
Vision and pattern recognition, 100–111
computational theory of vision, 105–108
feature detection theory, 103–105
feature integration theory, 108–111
stage theory, 105
template matching theory, 100–102
Visual agnosias, 176–181
apperceptive agnosia, 177–178
associative agnosia, 177–178
defined, 177
face perception, 180–181
Visual buffer, 141
Visual code, 131
Visual image, 4–5, 139
Visual imagery, 139–147
image processes, 142–147
image structures, 141–142
Visual object recognition, 175–181
Visual stage, 108
Visuo-spatial sketchpad, 138
Volitional consciousness, 44
Voluntarism movement, 68–72

w, 46
Wason Selection Task, 251–252
Weights, 211–212, 220
Wernicke-Geschwind Model, 294–298
Wernicke's aphasia, 295
Wernicke's area, 295
What-it's-like argument, 50–52
Whole-report condition, 127
Williams Syndrome, 99
Word recognition, logogen model of, 304–308
Working memory, 127–133
artificial intelligence approach, 438–439
capacity of, 129
cognitive approach, 435–436
components of, 136–138
evaluating neuroscience of, 195
evolutionary approach, 438
integrated program of study for, 434–439
linguistic approach, 438
model of, 137–139
network approach, 437
neural substrates of, 193–197
neuroscience approach, 437
philosophical approach, 435
psychological approach, 435
robotics approach, 439
search in, 159–160

Yerkes-Dodson law, 117

Zeigarnik effect, 91

About the Authors

Jay Friedenberg is Associate Chair and Associate Professor of the Psychology Department at Manhattan College, where he directs the Cognitive Science program. He obtained his Ph.D. in cognitive psychology in 1995 at the University of Virginia. His academic interests are in the areas of vision and philosophy of mind. He teaches courses in introductory and physiological psychology, sensation and perception, and research methods.

Dr. Friedenberg has an active research program investigating the visual estimation of center of mass. He has additionally published articles on symmetry detection and face perception. He is a member of the New York Academy of Sciences, The Psychonomic Society, Phi Beta Kappa, and Sigma Xi. He is also a yoga instructor and lives in Riverdale, New York, with a fat cat called Mimi.

Gordon Silverman is Chair of Electrical and Computer Engineering at Manhattan College. His professional career spans more than 50 years of corporate, teaching, consulting, and research experience, during which he has developed a range of scientific instruments, particularly for use in physiological psychology research environments. He is the holder of eight patents, some related to behavior modification. The author of more than 20 journal articles and books, he has also served on the faculties of The Rockefeller University and Fairleigh Dickinson University.

Dr. Silverman's current research interests include telemedicine, rehabilitation medicine, artificial intelligence, and biomedical instrumentation and modeling.